GARP®

FRM® | Financial Risk Manager

2023

FRM®

EXAM PART II

Risk Management and Investment Management

Pearson

Grateful acknowledgment is made to the following sources for permission to reprint material copyrighted or controlled by them:

"Factor Theory," "Factors," "Alpha (and the Low-Risk Anomaly)" by Andrew Ang, reprinted from *Asset Management: A Systematic Approach to Factor Investing* (2014), by permission of Oxford University Press.

"Portfolio Construction," by Richard Gringold and Ronald Kahn, reprinted from *Active Porfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk, 2nd edition*, (2000), by permission of McGraw-Hill Companies.

"Empirical Tests of Biases in Equity Portfolio Optimization," by Peter Muller, in *Financial Optimization,* edited by Stavros A. Zenios (Cambridge: Cambridge University Press, 1993), Table 4–4.

"Portfolio Risk: Analytical Methods" and "VaR and Risk Budgeting in Investment Management" by Philippe Jorion, reprinted from *Value at Risk: The New Benchmark for Managing Financial Risk, 3rd edition*, (2007), by permission of McGraw-Hill Companies.

"Risk Monitoring and Performance Management," by Bob Litterman and the Quantitative Resources Group, reprinted from *Modern Investment Management: An Equilibrium Approach* (2003), by permission of John Wiley & Sons, Inc. All rights reserved. Used under license from John Wiley & Sons, Inc.

"Portfolio Performance Evaluation," by Zvi Bodie and Alan J. Marcus, reprinted from *Investments*, 12th edition (2020), by permission of the McGraw-Hill Companies.

"Hedge Funds" reproduced from William Fung and David A. Hsieh, *Handbook of the Economics of Finance, Volume 2B: Financial Markets and Asset Pricing*, 9780444535948, 2013, Constantinides et al. Ch 16, Pg 1063–1125.

"Performing Due Diligence on Specific Managers and Funds," by Kevin R. Mirabile, reprinted from *Hedge Fund Investing: A Practical Approach to Understanding Investor Motivation, Manager Profits, and Fund Performance,* Second Edition (2016), by permission of John Wiley & Sons, Inc. All rights reserved. Used under license from John Wiley & Sons, Inc.

"Predicting fraud by investment managers" is reprinted from the *Journal of Financial Economics*, Vol 105, Stephen G. Dimmock, William C. Gerken, "Predicting fraud by investment managers," 153–173, 2012, with permission from Elsevier.

Learning Objectives provided by the Global Association of Risk Professionals.

Pearson Education, Inc., 330 Hudson Street, New York, New York 10013
A Pearson Education Company
www.pearsoned.com

Printed in the United States of America

ScoutAutomatedPrintCode

00022446-00000005 / 9780138028053

EEB/MB

Pearson

ISBN 10: 0-138-02805-2
ISBN 13: 978-013-802805-3

Contents

PREFACE

On behalf of our Board of Trustees, GARP's staff, and particularly its certification and educational program teams, I would like to thank you for your interest in and support of our Financial Risk Manager (FRM®) program.

The past few years have been especially difficult for the financial-services industry and those working in it because of the many disruptions caused by COVID-19. In that regard, our sincere sympathies go out to anyone who was ill, suffered a loss due to the pandemic, or whose career aspirations or opportunities were hindered.

The FRM program has experienced many COVID-related challenges, but GARP has always placed candidate safety first. During the pandemic, we've implemented many proactive measures to ensure your safety, while meeting all COVID-related requirements issued by local and national authorities. For example, we cancelled our entire exam administration in May 2020, and closed testing sites in specific countries and cities due to local requirements throughout 2020 and 2021. To date in 2022, we've had to defer many FRM candidates as a result of COVID.

Whenever we were required to close a site or move an exam administration, we affirmatively sought to mitigate the impact on candidates by offering free deferrals and seeking to create additional opportunities to administer our examinations at different dates and times, all with an eye toward helping candidates work their way through the FRM program in a timely way.

It's been our objective to assure exam candidates we will do everything in our power to assist them in furthering their career objectives in what is and remains a very uncertain and trying professional environment. Although we could not control the effects of the pandemic on our exam administration, we were able to ensure that none of the changes resulted in additional charges to candidates. In addition to free deferrals, we provided candidates with new materials at no cost when those unavoidable deferrals rolled into a new exam period, in which new topics were covered due to curriculum updates.

Since its inception in 1997, the FRM program has been the global industry benchmark for risk-management professionals wanting to demonstrate objectively their knowledge of financial risk-management concepts and approaches. Having FRM holders on staff gives companies comfort that their risk-management professionals have achieved and demonstrated a globally recognized level of expertise.

Over the past few years, we've seen a major shift in how individuals and companies think about risk. Although credit and market risks remain major concerns, operational risk and resilience and liquidity have made their way forward to become areas of material study and analysis. And counterparty risk is now a bit more interesting given the challenges presented by a highly volatile and uncertain global environment.

The coming together of many different factors has changed and will continue to affect not only how risk management is practiced, but also the skills required to do so professionally and at a high level. Inflation, geopolitics, stress testing, automation, technology, machine learning, cyber risks, straight-through processing, the impact of climate risk and its governance structure, and people risk have all moved up the list of considerations that need to be embedded into the daily thought processes of any good risk manager. These require a shift in thinking and raise

questions and concerns about whether a firm's daily processes are really fine-tuned, or if its data and information flows are fully understood.

As can be readily seen, we're living in a world where risks are becoming more complex daily. The FRM program addresses these and other risks faced by both non-financial firms and those in the highly interconnected and sophisticated financial-services industry. Because its coverage is not static, but vibrant and forward looking, the FRM has become the global standard for financial risk professionals and the organizations for which they work.

The FRM curriculum is regularly reviewed by an oversight committee of highly qualified and experienced risk-management professionals from around the globe. These professionals include senior bank and consulting practitioners, government regulators, asset managers, insurance risk professionals, and academics. Their mission is to ensure the FRM program remains current and its content addresses not only standard credit and market risk issues, but also emerging issues and trends, ensuring FRM candidates are aware of not only what is important but also what we expect to be important in the near future.

We're committed to offering a program that reflects the dynamic and sophisticated nature of the risk-management profession.

We wish you the very best as you study for the FRM exams, and in your career as a risk-management professional.

Yours truly,

Richard Apostolik

President & CEO

FRM® COMMITTEE

Chairperson

Nick Strange, FCA
Senior Technical Advisor, Operational Risk & Resilience,
Prudential Regulation Authority, Bank of England

Members

Richard Apostolik
President and CEO, GARP

Richard Brandt
MD, Operational Risk Management, Citigroup

Julian Chen, FRM
SVP, FRM Program Manager, GARP

Chris Donohue, PhD
MD, GARP Benchmarking Initiative, GARP

Donald Edgar, FRM
MD, Risk & Quantitative Analysis, BlackRock

Hervé Geny
Former Group Head of Internal Audit, London Stock Exchange Group

Aparna Gupta
Professor of Quantitative Finance
A.W. Lawrence Professional Excellence Fellow
Co-Director and Site Director, NSF IUCRC CRAFT
Lally School of Management
Rensselaer Polytechnic Institute

John Hull
Senior Advisor
Maple Financial Professor of Derivatives and Risk Management,
Joseph L. Rotman School of Management, University of Toronto

Keith Isaac, FRM
VP, Capital Markets Risk Management, TD Bank Group

William May
SVP, Global Head of Certifications and Educational Programs, GARP

Attilio Meucci, PhD, CFA
Founder, ARPM

Victor Ng, PhD
Chairman, Audit and Risk Committee
MD, Head of Risk Architecture, Goldman Sachs

Matthew Pritsker, PhD
Senior Financial Economist and Policy Advisor/Supervision, Regulation, and Credit, Federal Reserve Bank of Boston

Samantha C. Roberts, PhD, FRM, SCR
Instructor and Consultant, Risk Modeling and Analytics

Til Schuermann, PhD
Partner, Oliver Wyman

Sverrir Þorvaldsson, PhD, FRM
Senior Quant, SEB

1

Factor Theory

Learning Objectives

After completing this reading you should be able to:

- Provide examples of factors that impact asset prices and explain the theory of factor risk premiums.
- Discuss the capital asset pricing model (CAPM) including its assumptions and explain how factor risk is addressed in the CAPM.
- Explain the implications of using the CAPM to value assets, including equilibrium and optimal holdings, exposure to factor risk, its treatment of diversification benefits, and shortcomings of the CAPM.
- Describe multifactor models and compare and contrast multifactor models to the CAPM.
- Explain how stochastic discount factors are created and apply them in the valuation of assets.
- Describe efficient market theory and explain how markets can be inefficient.

Excerpt is Chapter 6 of Asset Management: A Systematic Approach to Factor Investing, *by Andrew Ang.*

See bibliography on pp. 195–200.

1.1 CHAPTER SUMMARY

Assets earn risk premiums because they are exposed to underlying factor risks. The capital asset pricing model (CAPM), the first theory of factor risk, states that assets that crash when the market loses money are risky and therefore must reward their holders with high risk premiums. While the CAPM defines bad times as times of low market returns, multifactor models capture multiple definitions of bad times across many factors and states of nature.

1.2 THE 2008–2009 FINANCIAL CRISIS

During the financial crisis of 2008 and 2009, the price of most risky assets plunged. Table 1.1 shows that U.S. large cap equities returned −37%; international and emerging markets equities had even larger losses. The riskier fixed income securities, like corporate bonds, emerging market bonds, and high yield bonds, also fell, tumbling along with real estate. "Alternative" investments like hedge funds, which trumpeted their immunity to market disruptions, were no safe refuge: equity hedge funds and their fixed income counterparts fell approximately 20%. Commodities had losses exceeding 30%. The only assets to go up during 2008 were cash (U.S. Treasury bills) and safe-haven sovereign bonds, especially long-term U.S. Treasuries.

Why did so many asset classes crash all at once? And given that they did, was the concept of diversification dead?

In this chapter, we develop a theory of factor risk premiums. The factor risks constitute different flavors of *bad times* and the investors who bear these factor risks need to be compensated in equilibrium by earning factor risk premiums. Assets have risk premiums not because the assets themselves earn risk premiums; assets are bundles of factor risks, and it is the exposures to the underlying factor risks that earn risk premiums. These factor risks manifest during bad times such as the financial crisis in late 2008 and early 2009.

1.3 FACTOR THEORY

Factors are to assets what nutrients are to food. Table 1.2 is from the Food and Nutrition Board, which is part of the Institute of Medicine of the National Academies, and lists recommended intakes of the five macronutrients—water, carbohydrates, protein, fiber, and fat—for an "average" male, female, and child. Carbohydrates can be obtained from food made from cereals and grains. Protein is obtained from meat and dairy products. Fiber is available from wheat and rice. Fat we can consume from animals but also certain plant foods such as peanuts. Each type of food is a bundle of nutrients. Many foods contain more than just one macronutrient: for example, rice contains both

Table 1.1 Returns of Asset Classes in 2008

Cash	Three-month T-bill	1.3%
Core Bonds	Barcap Aggregate Index	5.2%
Global Bonds	Citigroup World Government	10.9%
TIPS	Citigroup US Inflation Linked	−1.2%
Emerging Market Bonds	JPM Emerging Markets Bond Index	−9.7%
US High Yield	Merrill Lynch High Yield Master	−26.3%
Large Cap Equity	S&P 500	−37.0%
Small Cap Equity	Russell 2000	−33.8%
International Equity	MSCI World ex US	−43.2%
Emerging Markets Equity	IFC Emerging Markets	−53.2%
Public Real Estate	NAREIT Equity REITS	−37.7%
Private Real Estate	NCREIF Property Index	−16.9%
Private Capital	Venture Economics (Venture and Buyouts)	−20.0%
Equity Hedge Funds	HFRI Equity Hedge Index	−20.6%
Fixed Income Hedge Funds	HFRI Fixed Income Index	−17.8%
Commodities	Dow Jones AIG Commodity Index	−35.7%

Table 1.2 Nutrients and Food

Macronutrients				
	Male	**Female**	**Child**	**Examples of Food**
Water	3.7 L/day	2.7 L/day	1.7 L/day	
Carbohydrates	130 g/day	130 g/day	130 g/day	Bread, Beans, Potato, Rice
Protein	56 g/day	46 g/day	19 g/day	Cheese, Milk, Fish, Soya bean
Fiber	38 g/day	25 g/day	25 g/day	Peas, Wheat, Rice
Fat	20–35% of calories		25–35% of calories	Oily fish, Peanuts, Animal fat

Source: Food and Nutrition Board, National Academies, 2004.

carbohydrates and fiber. Different individuals, whether sick or healthy, male or female, or young or old, have different macronutrient requirements. We eat food for the underlying nutrients; it is the nutrients that give sustenance.

Factor risks are the driving force behind assets' risk premiums. An important theory of factor risk is the CAPM, which we explore in the next section. The CAPM states that there is only one factor driving all asset returns, which is the market return in excess of T-bills. All assets have different exposures to the market factor and the greater the exposure, the higher the risk premium. The market is an example of a tradeable, investment factor. Other examples include interest rates, value-growth investing, low volatility investing, and momentum portfolios. Factors can also be fundamental macro factors, like inflation and economic growth. Assets have different payoffs during high or low inflation periods or during economic recessions and expansions. We leave a complete exposition of the various types of factors to the next chapter. In this chapter, we describe the underlying theory of factor risk.

There are three similarities between food and assets:

1. Factors matter, not assets.

   If an individual could obtain boring, tasteless nutrients made in a laboratory, she would comfortably meet her nutrient requirements and lead a healthy life. (She would, however, deprive herself of gastronomic enjoyment.) The factors behind the assets matter, not the assets themselves. Investing right requires looking through asset class labels to understand the factor content, just as eating right requires looking through food labels to understand the nutrient content.

2. Assets are bundles of factors.

   Foods contain various combinations of nutrients. Certain foods are nutrients themselves—like water—or are close to containing only one type of nutrient, as in the case of rice for carbohydrates. But generally foods contain many nutrients. Similarly, some asset classes can be considered factors themselves—like equities and government fixed income securities—while other assets contain many different factors. Corporate bonds, hedge funds, and private equity contain different amounts of equity risk, volatility risk, interest rate risk, and default risk. Factor theory predicts these assets have risk premiums that reflect their underlying factor risks.

3. Different investors need different risk factors.

   Just as different people have different nutrient needs, different investors have different optimal exposures to different sets of risk factors.

   Volatility, as we shall see, is an important factor. Many assets and strategies lose money during times of high volatility, such as observed during the 2007–2008 financial crisis. Most investors dislike these times and would prefer to be protected against large increases in volatility. A few brave investors can afford to take the opposite position; these investors can weather losses during bad times to collect a volatility premium during normal times. They are paid risk premiums as compensation for taking losses—sometimes big losses, as in 2008–2009—during volatile times.

   Another example is that investors have different desired exposure to economic growth. One investor may not like times of shrinking GDP growth because he is likely to become unemployed in such circumstances. Another investor—a bankruptcy lawyer, perhaps—can tolerate low GDP growth because his labor income increases during recessions. The point is that each investor has different preferences, or risk aversion coefficients, for each different source of factor risk.

There is one difference, however, between factors and nutrients. Nutrients are inherently good for you. Factor risks are bad. It is by enduring these bad experiences that we are rewarded with risk premiums. Each different factor defines a different set

of bad times. They can be bad economic times—like periods of high inflation and low economic growth. They can be bad times for investments—periods when the aggregate market or certain investment strategies perform badly. Investors exposed to losses during bad times are compensated by risk premiums in good times. The factor theory of investing specifies different types of underlying factor risk, where each different factor represents a different set of bad times or experiences. We describe the theory of factor risk by starting with the most basic factor risk premium theory—the CAPM, which specifies just one factor: the market portfolio.

1.4 CAPM

The CAPM was revolutionary because it was the first cogent theory to recognize that the risk of an asset was not how that asset behaved in isolation but how that asset moved in relation to other assets and to the market as a whole. Before the CAPM, risk was often thought to be an asset's own volatility. The CAPM said this was irrelevant and that the relevant measure of risk was how the asset covaried with the market portfolio—the beta of the asset. It turns out that asset volatility itself matters, as we shall see in Chapter 2, but for the purpose of describing the CAPM and its incredible implications, we can ignore this for the time being.

The CAPM was formulated in the 1960s by Jack Treynor (1961), William Sharpe (1964), John Lintner (1965), and Jan Mossin (1966), building on the principle of diversification and mean-variance utility introduced by Harry Markowitz in 1952. For their work on CAPM and portfolio choice, Sharpe and Markowitz received the 1990 Nobel Prize in economics. (Merton Miller was awarded the Nobel Prize the same year for contributions to corporate finance.) Lintner and Mossin, unfortunately, had both died by then. Treynor, whose original manuscript was never published, has never received the recognition that he deserved.

I state upfront that the CAPM is well known to be a spectacular failure. It predicts that asset risk premiums depend only on the asset's beta and there is only one factor that matters, the market portfolio. Both of these predictions have been demolished in thousands of empirical studies. But, the failure has been glorious, opening new vistas of understanding for asset owners who must hunt for risk premiums and manage risk.

The basic intuition of the CAPM still holds true: that the factors underlying the assets determine asset risk premiums and that these risk premiums are compensation for investors' losses during bad times. Risk is a property not of an asset in isolation but how the assets move in relation to each other. Even though the CAPM is firmly rejected by data, it remains the workhorse model of finance: 75% of finance professors advocate using it, and 75% of CFOs employ it in actual capital budgeting decisions despite the fact that the CAPM does not hold.[1] It works approximately, and well enough for most applications, but it fails miserably in certain situations (as the next chapter will detail). Part of the tenacious hold of the CAPM is the way that it conveys intuition of how risk is rewarded.

What does the CAPM get right?

CAPM Lesson 1: Don't Hold an Individual Asset, Hold the Factor

The CAPM states that one factor exists and that factor is the market portfolio, where each stock is held in proportion to its market capitalization. This corresponds to a market index fund. The factor can be optimally constructed by holding many assets so that nonfactor, or idiosyncratic risk, is diversified away. Asset owners are better off holding the factor—the market portfolio—than individual stocks. Individual stocks are exposed to the market factor, which carries the risk premium (it is the nutrient), but also have *idiosyncratic risk*, which is not rewarded by a risk premium (this is the part that carries no nutritional value). Investors can diversify away the idiosyncratic part and increase their returns by holding the market factor portfolio, rather than any other combination of individual stocks. The market portfolio represents *systematic risk*, and it is pervasive: all risky assets have risk premiums determined only by their exposure to the market portfolio. Market risk also affects all investors, except those who are infinitely risk averse and hold only risk-free assets.

The key to this result is diversification. The CAPM is based on investors having mean-variance utility and the most important concept in mean-variance investing is diversification. Diversification ensures that, absent perfect correlation, when one asset performs badly, some other assets will perform well, and so gains can partly offset losses. Investors never want to hold assets in isolation; they improve their risk-return trade-off by diversifying and holding portfolios of many assets. This balance across many assets that are not perfectly correlated improves Sharpe ratios. Investors will diversify more and more until they hold the most diversified portfolio possible—the market portfolio. The market factor is the best, most-well diversified portfolio investors can hold under the CAPM.

The CAPM states that the market portfolio is held by every investor—a strong implication that is outright rejected in data.

[1] See Welch (2008) and Graham and Harvey (2001), respectively.

Nevertheless, it is useful to understand how we can leap from a diversified portfolio to the market being the only relevant factor.

The mean-variance frontier with the capital allocation line (CAL), is shown in Figure 1.1. This is the solution to the mean-variance investing problem. Investors hold different amounts of the risk-free asset and the mean-variance efficient (MVE) portfolio depending on their risk aversion. Now here come the strong assumptions of the CAPM. Assume that the set of means, volatilities, and correlations are the same for all investors. Then all investors hold the same MVE portfolio—just in different quantities depending on their own risk aversion. Since everyone holds the same MVE and this is the best portfolio that can be held by all investors, the MVE portfolio becomes the market factor in equilibrium.

Equilibrium

The equilibrium concept is extremely important. Equilibrium occurs when investor demand for assets is exactly equal to supply. The market is the factor in equilibrium because in CAPM land, everyone holds the MVE portfolio (except for those who are infinitely risk averse). If everyone's optimal risky portfolio (which is the MVE) assigns zero weight to a certain asset, say AA stock, then this cannot be an equilibrium. Someone must hold AA so that supply equals demand. If no one wants to hold AA, then AA must be overpriced and the expected return of AA is too low. The price of AA falls. The expected payoff of AA stays constant under CAPM assumptions, so that as the price of AA falls, the expected return of AA increases. AA's price falls until investors want to hold exactly the number of AA shares outstanding. Then, the expected return is such that supply is equal to demand in equilibrium. Since all investors hold the MVE portfolio, the MVE portfolio becomes the market portfolio, and the market consists of each asset in terms of market capitalization weights.

Equilibrium ensures that the factor—the market portfolio—will have a risk premium and that this risk premium will not disappear. The market factor is *systematic* and affects all assets. The market risk premium is a function of the underlying investors' risk aversions and utilities. That is, the risk premium of the market factor reflects the full setup of all people in the economy. The factors that we introduce later—tradeable factors like value-growth investing and volatility investing or macro factors like inflation and economic growth—will also carry risk premiums based on investor characteristics, the asset universe, and the production capabilities of the economy. They will disappear only if the economy totally changes. Equilibrium factor risk premiums will not disappear because clever hedge funds go and trade them—these types of investment strategies are not factors. Investors cannot arbitrage away the market factor and all other systematic factors.

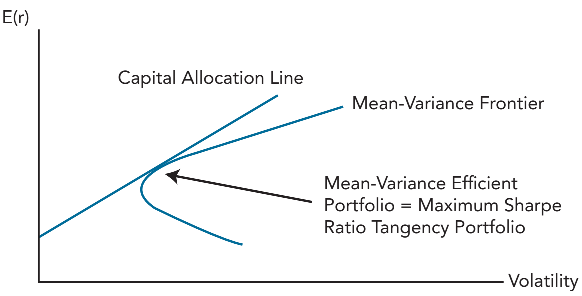


Figure 1.1

CAPM Lesson 2: Each Investor Has His Own Optimal Exposure of Factor Risk

In Figure 1.1, all investors will hold the market portfolio, just in different proportions. Pictorially, they have different proportions of the risk-free asset and the market portfolio and lie on different positions on the CAL line. Thus, each individual investor has a different amount of factor exposure just as different individuals have different nutrient requirements.

CAPM Lesson 3: The Average Investor Holds the Market

The market portfolio represents the average holdings across investors. The intersection of the CAL with the mean-variance frontier represents an investor who holds 100% in the MVE portfolio. This tangency point represents the average investor. The risk aversion corresponding to this 100% portfolio position is the risk aversion of the market.[2]

Note that as investors differ from the average investor, they will be exposed to more or less market factor risk depending on their own risk preferences.

CAPM Lesson 4: The Factor Risk Premium Has an Economic Story

The CAL in Figure 1.1 for a single investor is called the capital market line (CML) in equilibrium, since under the strong assumptions of the CAPM every investor has the same CML. (The MVE

[2] Technically speaking, the market is the wealth-weighted average across all investors.

portfolio is the market factor portfolio.) The equation for the CML pins down the risk premium of the market:

$$E(r_m) - r_f = \bar{\gamma}\sigma_m^2, \tag{1.1}$$

where $E(r_m) - r_f$ is the market risk premium, or the expected return on the market in excess of the risk-free rate; $\bar{\gamma}$ is the risk aversion of the "average" investor; and σ_m is the volatility of the market portfolio. The CAPM derives the risk premium in terms of underlying agent preferences ($\bar{\gamma}$ is the average risk aversion across all investors, where the average is taken weighting each individual's degree of risk aversion in proportion to the wealth of that agent).

According to the CAPM in Equation (1.1), as the market becomes more volatile, the expected return of the market increases and equity prices contemporaneously fall, all else equal. We experienced this in 2008 and 2009 when volatility skyrocketed and equity prices nosedived. Expected returns in this period on were very high (and realized returns were indeed high in 2009 and 2010). It is intuitive that the market risk premium in Equation (1.1) is proportional to market variance because under the CAPM investors have mean-variance preferences: they dislike variances and like expected returns. The market portfolio is the portfolio that has the lowest volatility among all portfolios that share the same mean as the market, or the market has the highest reward-to-risk ratio (or Sharpe ratio). The market removes all idiosyncratic risk. This remaining risk has to be rewarded, and Equation (1.1) states a precise equation for the risk premium of the market.

As the average investor becomes more risk averse to variance (so $\bar{\gamma}$ increases), the risk premium of the market also increases.

CAPM Lesson 5: Risk Is Factor Exposure

The risk of an individual asset is measured in terms of the factor exposure of that asset. If a factor has a positive risk premium, then the higher the exposure to that factor, the higher the expected return of that asset.

The second pricing relationship from the CAPM is the traditional beta pricing relationship, which is formally called the security market line (SML). Denoting stock *i*'s return as r_i and the risk-free return as r_f, the SML states that any stock's risk premium is proportional to the market risk premium:

$$\begin{aligned} E(r_i) - r_f &= \frac{\text{cov}(r_i, r_m)}{\text{var}(r_m)}(E(r_m) - r_f) \\ &= \text{b}_i(E(r_m) - r_f). \end{aligned} \tag{1.2}$$

The risk premium on an individual stock, $E(r_i) - r_f$, is a function of that stock's beta, $\beta_i = \text{cov}(r_i, r_m)/\text{var}(r_m)$. The beta is a measure of how that stock co-moves with the market portfolio, and the higher the co-movement (the higher $\text{cov}(r_i, r_m)$), the higher the asset's beta.

I will not formally derive Equation (1.2).[3] But it contains some nice intuition: mean-variance investing is all about diversification benefits. Beta—the CAPM's measure of risk—is a measure of *the lack of* diversification potential. Beta can be written as $\beta_i = \rho_{i,m}\sigma_i/\sigma_m$, where $\rho_{i,m}$ is the correlation between asset *i*'s return and the market return, σ_i is the volatility of the return of asset *i*, and σ_m is the volatility of the market factor. The lower the correlation with a portfolio, the greater the diversification benefit with respect to that portfolio because the asset was more likely to have high returns when the portfolio did badly. Thus, *high* betas mean *low* diversification benefits.

If we start from a diversified portfolio, investors find assets with high betas—those assets that tend to go up when the market goes up, and vice versa—to be unattractive. These high beta assets act like the diversified portfolio the investor already holds, and so they require high expected returns to be held by investors. In contrast, assets that pay off when the market tanks are valuable. These assets have low betas. Low beta assets have tremendous diversification benefits and are very attractive to hold. Investors, therefore, do not need to be compensated very much for holding them. In fact, if the payoffs of these low beta assets are high enough when the market return is low, investors are willing to pay to hold these assets rather than be paid. That is, assets with low enough betas actually have *negative* expected returns. These assets are so attractive because they have large payoffs when the market is crashing. Gold, or sometimes government bonds, are often presented as examples of low (or negative) beta assets which tend to pay off when the stock market crashes. (Government bonds were one of the few asset classes to do well in the financial crisis in 2008.)

CAPM Lesson 6: Assets Paying Off in Bad Times Have Low Risk Premiums

Another way to view the SML relationship in Equation (1.2) is that the risk premium in the CAPM is a reward for how an asset pays off in bad times. Bad times are defined in terms of the factor, which is the market portfolio, so bad times correspond to low (or negative) market returns. If the asset has losses when the market has losses, the asset has a high beta. When the market has gains, the high beta asset also gains in value. Investors are, on average, risk averse so that the gains during good times

[3] A textbook MBA treatment is Bodie, Kane, and Marcus (2011). A more rigorous treatment is Cvitanić and Zapatero (2004).

do not cancel out the losses during bad times. Thus, high beta assets are risky and require high expected returns to be held in equilibrium by investors.

Conversely, if the asset pays off when the market has losses, the asset has a low beta. This asset is attractive and the expected return on the asset can be low—investors do not need much compensation to hold these attractive assets in equilibrium. More generally, if the payoff of an asset tends to be high in bad times, this is a valuable asset to hold and its risk premium is low. If the payoff of an asset tends to be low in bad times, this is a risky asset and its risk premium must be high.

In the CAPM, the bad returns are defined as low returns of the market portfolio. This is, of course, very restrictive: there are many more factors than just the market. In multifactor models, all the intuitions of CAPM Lessons 1 through 6 hold. Except that, with multiple factors, each factor defines its own set of bad times.

1.5 MULTIFACTOR MODELS

Multifactor models recognize that bad times can be defined more broadly than just bad returns on the market portfolio. The first multifactor model was the *arbitrage pricing theory* (APT), developed by Stephen Ross (1976). It uses the word "arbitrage" because the factors cannot be arbitraged or diversified away—just like the single market factor in the CAPM. In equilibrium, investors must be compensated for bearing these multiple sources of factor risk. While the CAPM captures the notion of bad times solely by means of low returns of the market portfolio, each factor in a multifactor model provides its own definition of bad times.

Pricing Kernels

To capture the composite bad times over multiple factors, the new asset pricing approach uses the notion of a *pricing kernel*. This is also called a *stochastic discount factor* (SDF). We denote the SDF as m. The SDF is an index of bad times, and the bad times are indexed by many different factors and different states of nature. Since all the associated recent asset pricing theory uses this concept and terminology. It is worth spending a little time to see how this SDF approach is related to the traditional CAPM approach. There is some nice intuition that comes about from using the SDF, too. (For the less technically inclined, you are welcome to skip the next two subsections and start again at the Multifactor Model Lessons section.)

By capturing all bad times by a single variable m, we have an extremely powerful notation to capture multiple definitions of bad times with multiple variables. The CAPM is actually a special case where m is linear in the market return:[4]

$$m = a + b \times r_m, \tag{1.3}$$

for some constants a and b. (A pricing kernel that is linear in the market gives rise to a SML that with asset betas with respect to the market in Equation (1.2).) With our "m" notation, we can specify multiple factors very easily by having the SDP depend on a vector of factors, $F = (f_1, f_2, \ldots . f_K)$:

$$m = a + b_1 f_1 + b_2 f_2 + \ldots + b_K f_K, \tag{1.4}$$

where each of the K factors themselves define different bad times.

Another advantage of using the pricing kernel m is that while the CAPM restricts m to be *linear*, the world is *nonlinear*. We want to build models that capture this nonlinearity.[5] Researchers have developed some complicated forms for m, and some of the workhorse models describing equities and fixed income are nonlinear.

Pricing Kernels versus Discount Rate Models

Here's how these pricing kernels work. In the traditional *discount rate* model of the CAPM, we find the price of asset i by discounting its payoff next period back to today:

$$P_i = E\left[\frac{\text{payoff}}{1 + E(r_i)}\right], \tag{1.5}$$

where the discount rate is given by $E(r_i) = r_f + \beta_i(E(r_m) - r_f)$ according to the CAPM. Under the SDP model, we can equivalently write the price of the asset using m-notation:[6]

$$P_i = E[m \times \text{payoff}_i], \tag{1.6}$$

[4] The constants a and b can be derived using Equation (1.3) as

$$a = \frac{1}{1 + r_f} + \frac{\mu_m (\mu_m - r_f)}{(1 + r_f)\sigma_m^2},$$

$$b = -\frac{\mu_m - r_f}{(1 + r_f)\sigma_m^2},$$

where $\mu_m = E(r_m)$ and $\sigma_m^2 = \text{var}(r_m)$ are the mean and variance of the market returns, respectively. Note that the coefficient b multiplying m is negative: low values of the SDF correspond to bad times, which in the CAPM are given by low returns of the market.

[5] Related to this is that the requirement for the SDP is very weak: it requires only no arbitrage, as shown by Harrison and Kreps (1979). The CAPM, and other specific forms of m, on the other hand, require many additional onerous and often counterfactual assumptions.

[6] And, beyond the scope of this book, there are many useful statistical techniques for estimating m based on statistical "projections" similar to the estimation methods for ordinary least squares regressions based on the notation in Equation (1.4).

and hence the name "stochastic discount factor," because we discount the payoffs using m in Equation (1.6), just as we discount the payoff by the discount rate in the more traditional discount formula (1.5). The SDF is called a *pricing kernel*, borrowing the use of the word "kernel" from statistics, because one can estimate m in Equation (1.6) using a kernel estimator. Since it is a kernel that prices all assets, it is a "pricing kernel." Students of probability and statistics will recognize that the price in Equation (1.6) is an expectation taken with respect to the pricing kernel, so this gives rise to the SDF also being called the *state price density*.

We can divide both the right- and left-hand sides of Equation (1.6) by the asset's current price, P_i, to obtain

$$\frac{P_i}{P_i} = E\left[m \times \frac{\text{payoff}_i}{P_i}\right] \tag{1.7}$$
$$1 = E[m \times (1 + r_i)].$$

A special case of Equation (1.7) occurs when the payoffs are constant. That would give us a risk-free asset, so the price of a risk-free bond is simply $1/(1 + r_f) = E[m \times 1]$.

It turns out that we can write the risk premium of an asset in a relation very similar to the SML of the CAPM in Equation (1.2):[7]

$$E(r_i) - r_f = \frac{\text{cov}(r_i, m)}{\text{var}(m)}\left(-\frac{\text{var}(m)}{E(m)}\right) \tag{1.8}$$
$$= \beta_{i,m} \times \lambda_m,$$

where $\beta_{i,m} = \text{cov}(r_i, m)/\text{var}(m)$ is the beta of the asset with respect to the SDF. Equation (1.8) captures the "bad times" intuition that we had earlier from the CAPM. Remember that m is an index of bad times. The *higher* the payoff of the asset is in bad times (so the higher $\text{cov}(r_i, m)$ and the higher $\beta_{i,m}$), the *lower* the expected return of that asset. The higher beta in Equation (1.8) is multiplied by the price of "bad times" risk, $\lambda_m = -\text{var}(m)/E(m)$, which is the inverse of factor risk, which is why there is a negative sign. Equation (1.8) states directly the intuition of Lesson 6 from the CAPM: higher covariances with bad times lead to *lower* risk premiums. Assets that pay off in bad times are valuable to hold, so prices for these assets are high and expected returns are low.

Just as the CAPM gives rise to assets having betas with respect to the market, multiple factors in the SDF in Equation (1.4) gives rise to a multi-beta relation for an asset's risk premium:

$$E(r_i) = r_f + \beta_{i,1}E(f_1) + \beta_{i,2}E(f_2) + \ \ldots \ + \beta_{i,K}E(f_K), \tag{1.9}$$

where $\beta_{i,k}$ is the beta of asset i with respect to factor k and $E(f_k)$ is the risk premium of factor k. For macro factors, f_1 could be inflation and f_2 could be economic growth, for example.

[7] See, for example, Cochrane (2001) for a straightforward derivation,

Bad times are characterized by times of high inflation, low economic growth, or both. For an example for multiple investment factors, f_1 could be the market portfolio and f_2 could be an investing strategy based on going long value stocks and short growth stocks. Value stocks outperform growth stocks in the long run (see Chapter 2). Bad times are characterized by low market returns, value stocks underperforming growth stocks, or both.

Multifactor Model Lessons

The key lessons in the multifactor world are in fact the same from the CAPM:

	CAPM (Market Factor)	Multifactor Models
Lesson 1	Diversification works. The market diversifies away idiosyncratic risk.	Diversification works. The tradeable version of a factor diversifies away idiosyncratic risk.
Lesson 2	Each investor has her own optimal exposure of the market portfolio.	Each investor has her own optimal exposure of each factor risk.
Lesson 3	The average investor holds the market.	The average investor holds the market.
Lesson 4	The market factor is priced in equilibrium under the CAPM assumptions.	Risk premiums exist for each factor assuming no arbitrage or equilibrium.
Lesson 5	Risk of an asset is measured by the CAPM beta.	Risk of an asset is measured in terms of the factor exposures (factor betas) of that asset.
Lesson 6	Assets paying off in bad times when the market return is low are attractive, and these assets have low risk premiums.	Assets paying off in bad times are attractive, and these assets have low risk premiums.

The $64,000 question with multifactor pricing kernel models is: how do you define bad times? For the average investor who holds the market portfolio, the answer is when an extra $1 becomes very valuable. This interprets the SDF as the marginal utility of a *representative agent*. Times of high marginal utility are, for example, periods when you've just lost your job so your income is low and any extra dollars are precious to you. Your consumption is also low during these times. In terms of the average, representative consumer, this also corresponds to a macro factor definition of a bad time: bad times are when GDP growth is low, consumption is low, or economic growth in general is low. Times of high marginal utility could also be defined in relative terms: it could be when your consumption

is low relative to your neighbor or when your consumption is low relative to your past consumption. We captured the former using a catching up with the Joneses utility function and the latter with a habit utility function.

During 2008–2009, the financial crisis was a bad time with high volatility and large financial shocks. So volatility is an important factor, and the next chapter shows that many risky assets perform badly when volatility is high. Factors can also be tradeable, investment styles. Some of these include liquid, public market asset classes like bonds and listed equities. Others include investment styles that are easily replicable and that can be implemented cheaply (but often are not when they are delivered to customers) and in scale, like value/growth strategies.[8]

1.6 FAILURES OF THE CAPM

The CAPM is derived using some very strong assumptions. It's worth taking a moment to examine these assumptions and discuss what happens when they are relaxed.

1. Investors have only financial wealth.

   Investors have unique income streams and liabilities, and their optimal portfolio choice has to take these into consideration. Liabilities are often denominated in real terms—we want to maintain a standard of living even if prices rise, for example. Income streams are usually risky, and income declines during periods of low economic growth. This makes variables like inflation and growth important factors because many investors' income and liabilities change as the macro variables change.

   One particular important factor that drives asset returns is human capital, or labor income risk.[9] In an influential paper, Jagannathan and Wang (1996) found large improvements in the performance of the CAPM when labor income risk is taken into account.

2. Investors have mean-variance utility.

   More realistic utility functions often have an asymmetric treatment of risk because investors are generally more distressed by losses than pleased by gains. We should expect, then, to find deviations from the CAPM among stocks that have different measures of downside risk. Ang, Chen, and Xing (2006) show that stocks with greater downside risk have higher returns. A large number of papers show that other higher moment risk, like skewness and kurtosis, also carry risk premiums.[10]

3. Single-period investment horizon.

   By itself an investment horizon of one period is a minor assumption. Merton (1971, 1973) provides a famous extension of the CAPM to the dynamic case. In this setting, the CAPM holds in each single period.

   While the long investment horizon is an inconsequential assumption for the CAPM theory, there is a huge implication when we extend portfolio choice to a dynamic, long-horizon setting. The optimal strategy for long-term investors is to rebalance. The average investor, who holds the market portfolio by definition, does not rebalance.

4. Investors have *homogeneous* expectations.

   This assumption ensures that all investors hold the same MVE portfolio in the CAPM world and that, in equilibrium, the MVE portfolio becomes the market portfolio. In the real world, though, people obviously do not all share the same beliefs; they have *heterogeneous* expectations. By itself, the homogeneous expectations assumption is not important: a version of the CAPM holds where the expected returns are simply averages across the beliefs of all investors.[11] But, in combination with the next assumption, heterogeneous expectations can produce significant deviations from the CAPM.

5. No taxes or transactions costs.

   Taxes matter. Taxes affect expected returns and can be regarded as a systematic factor. Transactions costs, meanwhile, also vary across securities. We should expect that for very illiquid markets with high transactions costs, there may be more deviations from the CAPM. This is indeed the case, and Chapter 4 discusses various liquidity premiums in more detail.

   There is another effect of transaction costs when trading frictions are combined with heterogeneous investors. If investors cannot short, then investor beliefs matter. Optimists may

[8] There is a third type of factor based solely on statistical principal components, or similar (dynamic) statistical factor estimations of the APT. A pioneering example of these is Connor and Korajczyk (1986). These generally lack economic content, and so I do not discuss them here.

[9] Mayers (1973) is the seminal first reference. See also Constantinides and Duffie (1996), Jagannathan, Kubota, and Takehara (1998), Storesletten, Telmer, and Yaron (2007), and Eiling (2013).

[10] These effects come in two forms. First, there is the risk premium associated with individual stock higher moments. These are properties of each individual stock. See Mitton and Vorkink (2007), Boyer, Mitton, and Vorkink (2010), and Amaya et. al. (2012) for skewness risk premiums of this form. Second. there is the risk premium coming from how stock returns covary with higher moments of the aggregate market. Harvey and Siddique (2000), Dittmar (2002), and Chang, Christoffersen, and Jacobs (2013) show that there are risk premiums for co-skewness and co-kurtosis, which result from the co-movement of stock returns with skewness and kurtosis moments of the market portfolio.

[11] See Williams (1977).

prevail in pricing because the pessimists' beliefs are not impounded into stock prices. Pessimists would like to short but cannot, and so stock prices reflect only the belief of optimists. Thus, investor beliefs become a systematic factor. While there are behavioral versions of this story, the original setting of Miller (1977), where this concept was developed, was a rational setting. Related to this assumption is the next one, since when individuals move prices, markets are likely to be illiquid and there are many trading frictions.

6. Individual investors are price takers.

   The informed investor is trading and moving prices because he has some knowledge that others do not have. But when these trades are large, they move prices, which leads us to. . . .

7. Information is costless and available to all investors.

   Processing and collecting information is not costless, and certain information is not available to all investors. Information itself can be considered a factor in some economic settings, as in Veldkamp (2011). The CAPM applies in a stylized, efficient market; we should think that additional risk premiums can be collected in more inefficient securities markets, especially where information is very costly and not available to many investors. Several deviations from the CAPM are strongest in stocks that have small market capitalizations and trade in illiquid markets where information is not promulgated efficiently.

In summary, we expect that when the assumptions behind the CAPM are violated, additional risk premiums should manifest themselves. These include macro factors, which should affect investors' nonfinancial considerations, effects associated with the asymmetric treatment of risk, illiquidity and transactions costs, and taxes. We should expect failures of the CAPM to be most apparent in illiquid, inefficient markets. The assumption, in particular, of perfect information is one of the reasons why modern economists no longer believe that markets are efficient in the form the original CAPM specified.

1.7 THE FALL OF EFFICIENT MARKET THEORY

Today, economists do not believe in perfectly efficient markets,[12] In fact, markets cannot be efficient in their pure form. The modern notion of market near-efficiency is developed by Sanford Grossman and Joseph Stiglitz (1980), which forms part of the collection of papers for which Stiglitz was awarded his Nobel Prize in 2001. Grossman and Stiglitz describe a world in which markets are nearly efficient, and in doing so they address a conundrum that arises from the costless information assumption of the CAPM. Suppose that it is costly to collect information and to trade on that information, as it is in the real world. Then, if all information is in the price already, why would anyone ever invest in gathering the information? But if no one invests in gathering the information, how can information be reflected in security prices so that markets are efficient? It is then impossible that markets be efficient in their pure form.

Grossman and Stiglitz develop a model in which markets are near-efficient. Active managers search for pockets of inefficiency, and in doing so cause the market to be almost efficient. In these pockets of inefficiency, active managers earn excess returns as a reward for gathering and acting on costly information. In the assumptions of the CAPM discussed above, we should expect these pockets of inefficiency to lie in market segments that are illiquid, with poor information dissemination and where outsized profits may be hard to collect because trading on these anomalies will likely move prices.

The near-efficient market of Grossman and Stiglitz fits closely with the multiple factor risk framework of the APT developed by Ross (1976). In Ross's multifactor model, active managers and arbitrageurs drive the expected return of assets toward a value consistent with an equilibrium trade-off between risk and return. The factors in the APT model are systematic ones, or those that affect the whole economy, that agents wish to hedge against. In their purest form the factors represent risk that cannot be arbitraged away, and investors need to be compensated for bearing this risk.

Despite the modern notion that markets are not perfectly efficient, a large literature continues to test the *Efficient Market Hypothesis* (EMH). The implication of the EMH is that, to the extent that speculative trading is costly, active management is a loser's game and investors cannot beat the market.[13] The EMH does give us a very high benchmark: if we are average, we hold the market portfolio and indeed we come out ahead simply because we save on transactions costs. Even if we know the market cannot be perfectly efficient, tests of the EMH are still important because they allow investors to gauge where they may make excess returns. In the Grossman-Stiglitz context, talented investors can identify the pockets of inefficiency where active management efforts are best directed.

[12] The "classical" notions of weak, semi-strong, and strong efficiency were laid out by Fama (1970) and are obsolete. Fama was awarded the Nobel Prize in 2013. In that year, the Nobel Prize committee also gave Robert Shiller the prize, representing the opposite viewpoint of behavioral, or non-rational, influences on financial markets.

[13] Ellis (1975) for a practitioner perspective.

The EMH has been refined over the past several decades to rectify many of the original shortcomings of the CAPM including imperfect information and the costs associated with transactions, financing, and agency. Many behavioral biases have the same effect and some frictions are actually modeled as behavioral biases. A summary of EMH tests is given in Ang, Goetzmann, and Schaefer (2011). What is relevant for our discussion is that the deviations from efficiency have two forms: rational and behavioral. For an asset owner, deciding which prevails is important for deciding whether to invest in a particular pocket of inefficiency.

In a rational explanation, high returns compensate for losses during bad times. This is the pricing kernel approach to asset pricing. The key is defining those bad times and deciding whether these are actually bad times for an individual investor. Certain investors, for example, benefit from low economic growth even while the majority of investors find these to be bad periods. In a rational explanation, these risks premiums will not go away—unless there is a total regime change of the entire economy. (These are very rare, and the financial crisis in 2008 and 2009 was certainly not a regime change.) In addition, these risk premiums are scalable and suitable for very large asset owners.

In a behavioral explanation, high expected returns result from agents' under- or overreaction to news or events. Behavioral biases can also result from the inefficient updating of beliefs or ignoring some information. Perfectly rational investors, who are immune from these biases, should be able to come in with sufficient capital and remove this mispricing over time. Then it becomes a question of how fast an asset owner can invest before all others do the same. A better justification for investment, at least for slow-moving asset owners, is the persistence of a behavioral bias because there are barriers to the entry of capital. Some of these barriers may be structural, like the inability of certain investors to take advantage of this investment opportunity. Regulatory requirements, for example, force some investors to hold certain types of assets, like bonds above a certain credit rating or stocks with market capitalizations above a certain threshold. If there is a structural barrier to entry, then the behavioral bias can be exploited for a long time.

For some risk premiums, the most compelling explanations are rational (as with the volatility risk premium), for some behavioral (e.g., momentum), and for some others a combination of rational and behavioral stories prevails (like value/growth investing). Overall, the investor should not care if the source is rational or behavioral; the more appropriate question is whether she is different from the average investor who is subject to the rational or behavioral constraints and whether the source of returns is expected to persist in the future (at least in the short term).

1.8 THE 2008–2009 FINANCIAL CRISIS REDUX

The simultaneously dismal performance of many risky assets during the financial crisis is consistent with an underlying multifactor model in which many asset classes were exposed to the same factors. The financial crisis was the quintessential bad time: volatility was very high, economic growth toward the end of the crisis was low, and there was large uncertainty about government and monetary policy. Liquidity dried up in several markets. The commonality of returns in the face of these factor risks is strong evidence in favor of multifactor models of risk, rather than a rejection of financial risk theory as some critics have claimed. Assets earn risk premiums to compensate for exposure to these underlying risk factors. During bad times, asset returns are low when these factor risks manifest. Over the long run, asset risk premiums are high to compensate for the low returns during bad times.

Some commentators have argued that the events of 2008 demonstrate the failure of diversification. Diversification itself is not dead, but the financial crisis demonstrated that asset class labels can be highly misleading, lulling investors into the belief that they are safely diversified when in fact they aren't. What matters are the embedded factor risks; assets are bundles of factor risks. We need to understand the factor risks behind assets, just as we look past the names and flavors of the things that we eat to the underlying nutrients to ensure we have enough to sustain us. We take on risk to earn risk premiums in the long run, so we need to understand when and how that factor risk can be realized in the short run. Some have criticized the implementation of diversification through mean-variance utility, which assumes correlations between asset classes are constant when in fact correlations tend to increase during bad times.[14] Factor exposures can and do vary through time, giving rise to time-varying correlations—all the more reason to understand the true factor drivers of risk premiums.

[14] Models of portfolio choice with time-varying correlations are developed by Ang and Bekaert (2002, 2004), for example, Chua, Kritzman, and Page (2009) provide an analysis of increasing correlations during the financial crisis.

2

Factors

Learning Objectives

After completing this reading you should be able to:

- Describe the process of value investing and explain why a value premium may exist.
- Explain how different macroeconomic risk factors, including economic growth, inflation, and volatility, affect asset returns and risk premiums.
- Assess methods of mitigating volatility risk in a portfolio and describe challenges that arise when managing volatility risk.
- Explain how dynamic risk factors can be used in a multifactor model of asset returns, using the Fama-French model as an example.
- Compare value and momentum investment strategies, including their return and risk profiles.

Excerpt is Chapter 7 of Asset Management: A Systematic Approach to Factor Investing, *by Andrew Ang. See bibliography on pp. 195–200.*

2.1 CHAPTER SUMMARY

Factors drive risk premiums. One set of factors describes fundamental, economy-wide variables like growth, inflation, volatility, productivity, and demographic risk. Another set consists of tradeable investment styles like the market portfolio, value-growth investing, and momentum investing. The economic theory behind factors can be either rational, where the factors have high returns over the long run to compensate for their low returns during bad times, or behavioral, where factor risk premiums result from the behavior of agents that is not arbitraged away.

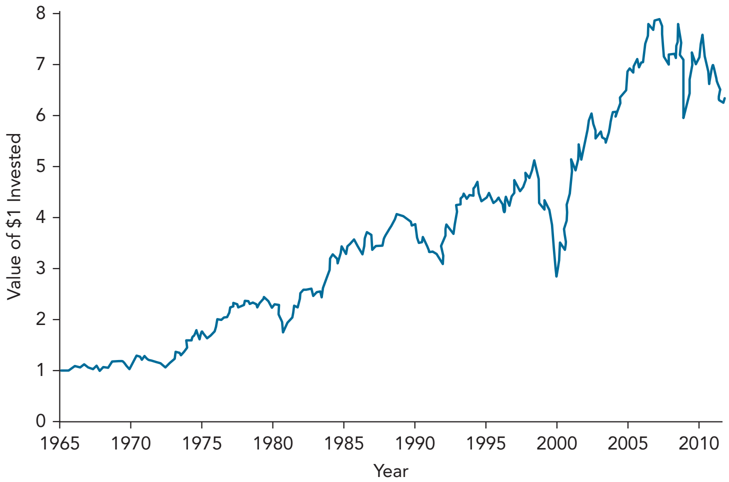


Figure 2.1 **Returns to the value-growth strategy.**

2.2 VALUE INVESTING

Historically speaking, value stocks beat the pants off growth stocks. Value stocks have low prices in relation to their net worth, which can be measured by accounting book value. Growth stocks are relatively costly in comparison to book value. Figure 2.1 plots the returns of value stocks (stocks with high book-to-market ratios) versus growth stocks (stocks with low book-to-market ratios). I plot the returns to the *value-growth strategy*, which goes long value stocks and short growth stocks.[1] Although value investing has on average done well, it sometimes loses money. For example, note the pronounced drawdown during the tech boom of the late 1990s. There was another drawdown during the financial crisis in 2008. Value stocks also did poorly in 2011.

Why does value investing work? Was the value strategy—the returns of value stocks in excess of growth stocks—a systematic factor? If so, what determined the value risk premium?

In the context of the previous chapter on factor theory, assets are buffeted by risk factors. The risk factors offer premiums to compensate investors for bearing losses during bad times. I discuss the economic stories behind the factors from a rational and behavioral perspective and the implications of these stories for asset owners.[2]

There are two types of factors. There are macro, fundamental-based factors, which include economic growth, inflation, volatility, productivity, and demographic risk. The second type is investment-style factors like the market factor of the capital asset pricing model (CAPM) and the value strategy of this motivating example. Investment factors include both *static factors*, like the market, which we simply go long to collect a risk premium, and *dynamic factors*, which can only be exploited through constantly trading different types of securities. Many hedge funds and private equity investments are essentially bundles of dynamic factors. The two types of factors are linked, and macro factors are often embedded in the performance of investment factors. I turn to economy-wide macro factors first.

2.3 MACRO FACTORS

It is intuitive that macro factors pervasively affect all investors and the prices of assets.[3] When economic growth slows or inflation is high, all firms and investors in the economy are affected—it is just a question of degree. Most consumers dislike low growth and high inflation because it is more likely they will be laid off or they are less able to afford the same basket of goods and services in real terms. A few investors, such as debt collectors, benefit from slow growth, and a few other investors, including owners of oil wells, benefit from high inflation induced by surging commodity prices. In general, though, bad outcomes of macro factors define bad times for the average investor.

The *level* of the factor often does not matter as much as a *shock* to a factor. Many macro factors are persistent: when inflation is

[1] The data for this strategy, as for all the other Fama-French strategies in this chapter are from http://mba.tuck.dartmouth.edu/ pages/facultylken .french/data_library.html.

[2] A very comprehensive study of factor risks is Ilmanen (2011).

[3] The first study to consider macro factors as systematic sources of risk in the cross section of equities was Chen, Roil, and Ross (1986).

00011693-00000002_CH02_p013-028.indd 14 05/07/22 3:12 PM

low today, we know that it will be very likely low next month. The fact that it is then low at the end of the month is no surprise. What is surprising are any movements in inflation not anticipated at the beginning of the period. Thus, we often need to look at *unexpected changes* to macro factors.

Asset prices respond to these factors *contemporaneously*. As inflation is increasing or unexpected adverse inflation shocks hit the economy, we enter a bad time and asset prices fall. The risk premium over the long run compensates investors for the Josses endured when bad times of high inflation occur in the short run.

The three most important macro factors are growth, inflation, and volatility.

Economic Growth

Risky assets generally perform poorly and are much more volatile during periods of low economic growth. However, government bonds tend to do well during these times. If an investor is in a position to weather recessions relatively comfortably, then that person should tilt more heavily toward risky assets such as equities. In doing so she'll enjoy higher returns, on average, and over the long run these will make up for losses during periods of low growth.[4] If an investor cannot bear large losses during recessions, she should hold more bonds, especially government bonds. Her portfolio will likely perform much better during recessions but worse over the long run. This is the price the investor pays for low exposure to growth risk.

Table 2.1 reports means and volatilities of large stocks, small stocks, government bonds, and corporate bonds (investment grade and high yield) conditional on economic recessions and expansions defined by the National Bureau of Economic Research (NBER). I also report means and volatilities conditional on low and high real GDP growth and low and high consumption. These are computed simply by dividing the sample into two sets, above and below the median, respectively. Table 2.1 shows that, during recessions, stock returns fall: the mean return for large stocks is 5.6% during recessions and 12.4% during expansions. The difference in returns across recessions and expansions is more pronounced for the riskier small cap stocks at 7.8% and 16.8%, respectively. Government bonds act in the opposite way, generating higher returns at 12.3% during recessions compared to 5.9% during expansions. Investment-grade corporate bonds, which have relatively little credit risk, exhibit similar behavior. In contrast, high-yield bonds are much closer to equity, and their performance is between equity and government bonds; in fact, high-yield bonds do not have any discernable difference in mean returns over recessions and expansions.

We can see a similar pattern if we look at periods of low or high growth, as measured by real GOP or consumption growth. For example, large stocks return 8.8% during periods of low real GDP growth and 5.6% during periods of low consumption growth. During periods of high real GDP growth and high consumption growth. large stock returns average 13.8% and 17.1%, respectively. Consistent with the behavior across NBER recessions and expansions, government bonds tend to do relatively well during periods of low growth, averaging 10.0% during periods of low real GDP growth compared to 3.9% during periods of high real GDP growth.

All asset returns are much more volatile during recessions or periods of low growth. For example, large stock return volatility is 23.7% during recessions compared to 14.0% during expansions. While government bonds have higher returns during recessions, their returns are also more volatile then, with a volatility of 15.5% during recessions compared to 9.3% during expansions. It is interesting to compare the volatilities of assets over the full sample to the volatilities conditional on recessions and expansions: volatility rends to be very high during bad times.

Inflation

High inflation tends to be bad for both stocks and bonds, as Table 2.1 shows. During periods of high inflation, all assets tend to do poorly.[5] Large stocks average 14.7% during low inflation periods and only 8.0% during periods of high inflation. The numbers for government bonds, investment grade bonds, and high yield bonds are 8.6%, 8.8%, and 9.2%, respectively, during low inflation periods and 5.4%, 5.3%, and 6.0%, respectively, during high inflation periods. It is no surprise that high inflation hurts the value of bonds: these are instruments with fixed payments, and high inflation lowers their value in real terms. It is more surprising that stocks—which are real in the sense that they represent ownership of real, productive firms—do poorly when inflation is high. High inflation is bad for both equities and bonds. Part of the long-run risk premiums for both equities and bonds represents compensation for doing badly when inflation is high.

[4] A related variable to GDP growth is real consumption growth. It turns out that real consumption is very smooth and actually does not vary much across recessions and expansions, unlike GOP growth.

[5] Deflation, which is not examined here, is also a bad time when assets tend to have low returns.

Table 2.1 Means and Volatilities Conditional on Factor Realizations

	Large Stocks	Small Stocks	Govt Bonds	Corporate Bonds: Investment Grade	Corporate Bonds: High Yield
Means					
Full Sample	11.3%	15.3%	7.0%	7.0%	7.6%
Business Cycles (1)					
Recessions	5.6%	7.8%	12.3%	12.6%	7.4%
Expansions	12.4%	16.8%	5.9%	6.0%	7.7%
Real GDP (2)					
Low	8.8%	12.2%	10.0%	9.7%	7.0%
High	13.8%	18.4%	3.9%	4.4%	8.2%
Consumption (3)					
Low	5.6%	5.6%	9.6%	9.1%	7.1%
High	17.1%	25.0%	4.4%	5.0%	8.2%
Inflation (4)					
Low	14.7%	17.6%	8.6%	8.8%	9.2%
High	8.0%	13.0%	5.4%	5.3%	6.0%
Volatilities					
Full Sample	16.0%	23.7%	10.6%	9.8%	9.5%
Business Cycles					
Recessions	23.7%	33.8%	15.5%	16.6%	18.1%
Expansions	14.0%	21.2%	9.3%	7.8%	6.8%
Real GDP					
Low	16.9%	23.7%	12.2%	11.8%	12.1%
High	14.9%	23.7%	8.5%	7.0%	6.0%
Consumption					
Low	17.5%	23.8%	11.9%	11.6%	11.8%
High	13.8%	22.7%	8.9%	7.4%	6.6%
Inflation					
Low	15.5%	21.9%	9.6%	8.2%	7.7%
High	16.4%	25.4%	11.5%	11.1%	11.0%

Returns are from Ibbotson Morningstar and are at the quarterly frequency.
The sample is 1952:Q1 to 2011:Q4.
(1) Business cycles are defined by NBER recession and expansion indicators.
(2) Real GDP is quarter-on-quarter.
(3) Consumption is quarter-on-quarter real personal consumption expenditures.
(4) Inflation is quarter-on-quarter CPI-All Items.

Volatility

Volatility is an extremely important risk factor. I measure volatility risk using the VIX index, which represents equity market volatility.[6] Here's a table correlating changes in VIX with stock and bonds returns on a monthly frequency basis from March 1986 to December 2011:

	VIX Changes	Stock Returns	Bond Returns
VIX Changes	1.00	−0.39	0.12
Stock Returns	−0.39	1.00	−0.01
Bond Returns	0.12	−0.01	1.00

The correlation between VIX changes and stock returns is −39%, so stocks do badly when volatility is rising. The negative relation between volatility and returns is called the *leverage effect*.[7] When stock returns drop, the financial leverage of firms increases since debt is approximately constant while the market value of equity has fallen. This makes equities riskier and increases their volatilities. There is another channel where high volatilities lead to low stock returns: an increase in volatility raises the required return on equity demanded by investors, also leading to a decline in stock prices. This second channel is a time-varying risk premium story and is the one that the basic CAPM advocates: as market volatility increases, discount rates increase and stock prices must decline today so that future stock returns can be high.[8]

Bonds offer some but not much respite during periods of high volatility, as the correlation between bond returns and VIX changes is only 0.12. Thus, bonds are not always a safe haven when volatility shocks hit. In 2008 and 2009, volatility was one of the main factors causing many risky assets to fall simultaneously. During this period, risk-free bonds did very well. But during the economic turbulence of the late 1970s and early 1980s, bonds did terribly, as did equities. Volatility as measured by VIX can also capture uncertainty—in the sense that investors did not know the policy responses that government would take during the financial crisis, whether markets would continue functioning, or whether their own models were the correct ones.

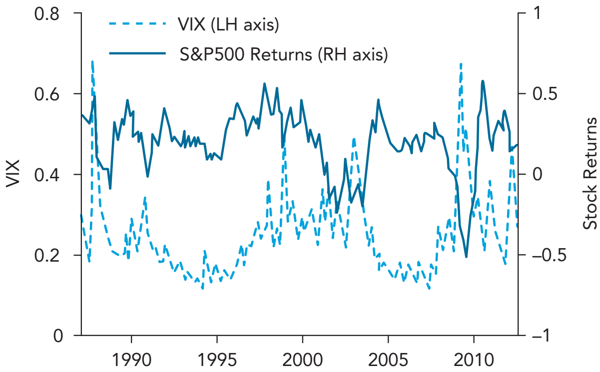


Figure 2.2 **VIX and one-year moving average of stock returns.**

Recent research posits uncertainty risk itself as a separate factor from volatility risk, but uncertainty risk and volatility risk are highly correlated.[9]

Figure 2.2 plots the VIX index (left-hand side axis) in the dashed line and a one-year moving average of stock returns (on the right-hand side axis) in the solid line. Volatility tends to exhibit periods of calm, punctuated by periods of turbulence. Figure 2.2 shows spikes in volatility corresponding to the 1987 stock market crash, the early 1990's recession, the Russian default crisis in 1998, the terrorist attacks and ensuing recession in 2001, and a large spike in 2008 corresponding to the failure of Lehman Brothers. In all of these episodes, stock returns move in the opposite direction from volatility, as shown during the financial crisis.

The losses when volatility spikes to high levels can be quite severe. Partitioning the sample into high and low periods of volatility changes gives us an average return for large stocks of –4.6% during volatile times and 24.9% during stable times. This compares to an overall mean of 11.3% (see Table 2.1). Investors allergic to volatility could increase their holdings of bonds, but bonds do not always pay off during highly volatile periods—as the low correlation of 0.12 between VIX changes and bond returns in Figure 2.2 shows.

Stocks are not the only assets to do badly when volatility increases. Volatility is negatively linked to the returns of many assets and strategies. Currency strategies fare especially poorly in times of high volatility.[10] We shall see later that many assets or strategies implicitly have large exposure to volatility risk. In particular, hedge funds, in aggregate, sell volatility.

[6] The VIX index is a measure of option volatilities on the S&P 500 constructed by the Chicago Board Options Exchange. The VIX index captures a variety of risks related to higher movements, including volatility itself, but also jump risk and skewness risk. But the main components captured in the VIX index are volatility and the volatility risk premium.

[7] A term coined by Fischer Black (1976).

[8] See Equation (1.2) in Chapter 1. For evidence of this time-varying risk premium channel, see Bekaert and Wu (2000).

[9] See, for example, Anderson, Ghysels, and Juergens (2009) for stocks and Ulrich (2011) for bonds.

[10] See Bhansali (2007) and Menkhoff et. al. (2012a).

Investors who dislike volatility risk can buy volatility protection (e.g., by buying put options). However, some investors can afford to take on volatility risk by selling volatility protection (again, e.g., in the form of selling put options). Buying or selling volatility protection can be done in option markets, but traders can also use other derivatives contracts, such as volatility swaps. Investors are so concerned about volatility, on average, that they are willing to pay to avoid volatility risk, rather than be paid to take it on. Periods of high volatility coincide with large downward movements (see Figure 2.2) and assets that pay off during high volatility periods, like out-of-the-money puts, provide hedges against volatility risk.

We often think about assets having *positive* premiums—we buy, or go long, equities, and the long position produces a positive expected return over time. Volatility is a factor with a *negative* price of risk. To collect a volatility premium requires *selling* volatility protection, especially selling out-of-the-money put options. The VIX index trades, on average, above volatilities observed in actual stocks: VIX implied volatilities are approximately 2% to 3%, on average, higher than realized volatilities. Options are thus expensive, on average, and investors can collect the volatility premium by short volatility strategies. Fixed income, currency, and commodity markets, like the aggregate equity market, have a negative price of volatility risk.[11]

Selling volatility is not a free lunch, however. It produces high and steady payoffs during stable times. Then, once every decade or so, there is a huge crash where sellers of volatility experience large, negative payoffs. Figure 2.3 plots the cumulated returns of a volatility premium (swap) index constructed by Merrill Lynch. There are steady returns until 2007, with the few blips corresponding to some small losses during 1998 (Russian default crisis) and 2001 and 2002 (9/11 tragedy and economic recession, respectively), and also during the summer of 2007 (subprime mortgage-backed losses just prior to the financial crisis). But between September and November 2008 there are massive losses close to 70%. These were the darkest months in the financial crisis, and

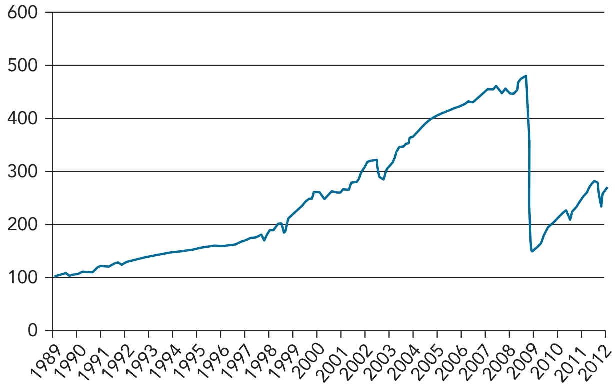


Figure 2.3 Cumulated returns from selling volatility.

most of the losses across all types of risky assets during 2008 were concentrated during these months (see Table 1.1). The huge crash causes volatility selling to have a large negative skewness of –8.26 over the whole sample. Taking data prior to the financial crisis ending December 2007, the skewness is only –0.37, so prior to 2007 selling volatility looked like easy money.

Unfortunately, some investors who sold volatility prior to the financial crisis failed to anticipate that a crash like the one of 2008 would materialize. But it did, and the abysmal returns of many assets resulted to a great extent from them being exposed to volatility risk. While forecasting when a crash will take place is always hard, if not impossible, investors might have known from past data that a crash of this type will happen from time to time. For example, volatility had spiked to these levels during the Great Depression. In fact, during the 1930s, volatility was not only extremely high, but it remained high for a much longer period than the 2008–2009 experience.

Only investors who can tolerate periods of very high volatility—which tend to coincide with negative returns on most risky assets—should be selling volatility protection through derivatives markets. Selling volatility is like selling insurance. During normal times, you collect a premium for withstanding the inevitable large losses that occur every decade or so. The losses endured when volatility spikes represent insurance payouts to investors who purchased volatility protection.

Rebalancing as a portfolio strategy is actually a short volatility strategy. Thus, the simple act of rebalancing will reap a long-run volatility risk premium, and the person who does not rebalance—the average investor who owns 100% of the market—is long volatility risk and loses the long-run volatility risk premium.[12] A long-run, rebalancing investor is exposed to the possibilities of fat,

[11] For a negative volatility risk premium in fixed income markets, see Simon (2010) and Mueller. Vedolin, and Yen (2012); currency markets, see Low and Zhang (2005); commodity markets, see Prokopczuk and Wese (2012); and the aggregate stock market, see Bakshi and Kapadia (2003) and Ang et al. (2006). For individual stocks, the volatility risk premium can be positive (some agents really like individual stock risk), see Driessen, Macnhout, and Vilkov (2009). One explanation why individual stocks can carry positive risk premiums but the volatility risk premium is significantly negative at the aggregate level is that much of a stock's variation is idiosyncratic. In portfolios, the stock-specific, idiosyncratic movements are diversified away leaving only market volatility risk, which has a negative risk premium. In the discussion in the main text, volatility risk refers to both "smooth" movements in time-varying volatility (*diffusive* risk) and abrupt changes (*jump* risk). More sophisticated models differentiate between the two; see Pan (2002).

[12] Sharpe (2010) calls a non-rebalancing strategy an adaptive allocation policy.

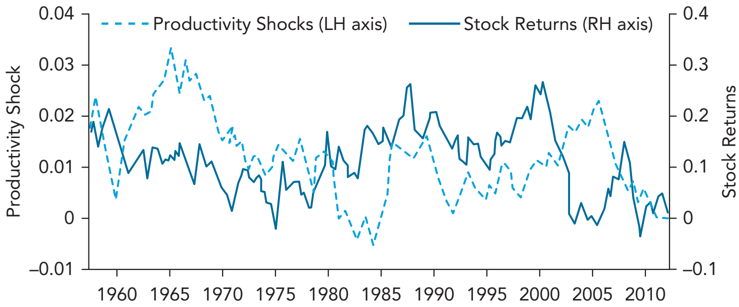


Figure 2.4 **5-Year moving average of productivity and stock returns, corr = 0.48.**

left-hand tail losses like those in Figure 2.3. There are two differences, however. Rebalancing over assets does not directly trade volatility risk. That is, rebalancing over stocks trades *physical* stocks, but Figure 2.3 involves trading *risk-neutral*, or option, volatility. Trading volatility in derivatives markets brings an additional volatility risk premium that rebalancing does not. Thus, losses in trading volatility in derivative markets are potentially much steeper than simple rebalancing strategies. Second, pure volatility trading in derivatives can be done without taking any stances on expected returns through delta-hedging. Rebalancing over fundamental asset or strategy positions is done to earn underlying factor risk premiums. While there is only weak predictability of returns, the investor practicing rebalancing gets a further boost from mean reversion as she buys assets with low prices that have high expected returns.

Constructing valuation models with volatility risk can be tricky because the relation between volatility and expected returns is time varying and switches signs and is thus *very* hard to pin down. A large literature has tried to estimate the return-volatility trade-off as represented in Equation (1.1) repeated here:

$$E(r_m) - r_f = \bar{\gamma}\sigma_m^2, \tag{2.1}$$

where $E(r_m) - r_f$ is the market risk premium and σ_m^2 is the variance of the market return. According to CAPM theory, $\bar{\gamma}$ represents the risk aversion of the average investor.

Is the coefficient, $\bar{\gamma}$, relating the market volatility or variance to expected returns, which is supposedly positive in theory, actually positive in data? In the literature, there are estimates that are positive, negative, or zero. In fact, one of the seminal studies, Glosten, Jagannathan, and Runkle (1993), contains all three estimates in the same paper! Theoretical work shows that the risk–return relation can indeed be negative and change over time.[13] What is undisputed, though, is that when volatility increases dramatically, assets tend to produce losses. Only an investor who can tolerate large losses during high-volatility periods should consider selling volatility protection.

Other Macro Factors

Several other macro factors have been investigated extensively in the literature and deserve attention from asset owners.

Productivity Risk

A class of *real business cycle models* developed in macroeconomics seeks to explain the movements of macro variables (like growth, investment, and savings) and asset prices across the business cycle. In these models, macro variables and asset prices vary across the business cycle as a rational response of firms and agents adjusting to *real shocks*. The label "real" in "real business cycle" emphasizes that the business cycle is caused by real shocks and is not due to market failures or insufficient demand as in the models of John Maynard Keynes (1936). Real business cycle models have inflation, but inflation is *neutral* or has no real effects. These models are *production economies* because they involve optimizing firms producing physical goods, in addition to agents optimizing consumption and savings decisions, but the firms are subject to shocks that affect their output. One particularly important shock that affects firm output is a *productivity shock*. The early literature, like Kydland and Prescott (1982), did not have asset prices. (Kydland and Prescott won the Nobel Prize in 2004.) The next generation of models, like Jermann (1988), put them in. The newest papers, like Kaltenbrunner and Lochstoer (2010), capture realistic, and complicated, dynamics of shocks and agents' behavior.

Because these models are designed to work at business cycle frequencies, they are less relevant for investors who have short horizons. But for long-horizon investors—like certain pension funds, sovereign wealth funds, and family offices—the productivity factor should be considered. Asset returns reflect long-run productivity risk. At the time of writing, Europe is still in the throes of its sovereign debt convulsions, and an important issue is the future productive capacity of European economies. In Figure 2.4, I plot a five-year average of productivity shocks and stock returns. I use a five-year average because the productivity variable is used in economic models that are designed to explain business cycle variation, which has a frequency of three to six years. The productivity shocks are alternatively called *Solow residuals* after Robert Solow (1957) or *total factor productivity* (TFP) shocks. I take the TFP shocks constructed by Fernald (2009), who follows the method of Basu, Fernald, and Kimball

[13] See, for example, Backus and Gregory (1993), Whitelaw (2000), and Ang and Liu (2007).

(2006).[14] Figure 2.4 shows that when there are periods of falling productivity, like the 1960s and 1970s, stock prices tend to fall. In the 1980s and 1990s (the computer revolution), productivity shocks are positive and stocks tend to appreciate. The correlation of the five-year moving averages of TFP shocks and stock returns is high, at 48%. So stocks are exposed to productivity risk; when productivity slows down, stock returns tend to be low.

Productivity risk is just one source of shocks that enter the new generation of *dynamic stochastic general equilibrium* (DSGE) macro models. This mouthful conveys the complexity of this class of models. In DSGE models, the economy is dynamic (as the name indicates) and the actions of agents (consumers, firms, ccntral banks, and governments), technologies (how firms produce), and institutions or markets (the way that agents interact) cause economic variables to change. Asset prices are set from the complex interaction of all of these players and technologies. The DSGE models allow us to think about how shocks from these factor risks are transmitted across the economy. An important part of DSGE models are the actions of policy makers—government policy matters, as the financial crisis showed. Monetary policy and government shocks are important factors that influence asset prices and constitute their own sources of risk. Current DSGE models nest both the real business cycle models pioneered by Kydland and Prescott, and they also include *new-Keynesian models*, where prices do not immediately adjust and inflation is *non-neutral*.

DSGE models describe business cycle fluctuations well, and we know asset returns vary over the business cycle. A benchmark model today is Smets and Wouters (2007), who specify seven shocks: productivity (as we have just discussed), investment, preferences, labor supply, inflation, government spending, and monetary policy.

Demographic Risk

Another important risk for a very long-term investor is *demographic risk*. This can be interpreted as a shock to labor output, just as a productivity shock is a shock to firm production. A slow-moving variable, demography is a factor in economic *overlapping generations* (OLG) models. A given individual follows a life-cycle model. Take, for example, an individual who progresses through youth, middle-age, and retirement. Labor income is earned and saved only during the young and middle-aged periods, and dis-saving occurs when retired. As any given age cohort progresses through the three stages, they join two other cohorts already alive who were born in previous generations. Thus, several generations overlap at any given time. A demographic shock changes the composition of a given cohort relative to other cohorts through such events as war (like World Wars I and II), a baby boom (like the generation born in the two decades following World War II), or infectious disease (Spanish Flu in 1918).

Several OLG models predict that demographic composition affects expected returns. Theory suggests two main avenues for this to occur. First, the life-cycle smoothing in the OLG framework requires that when the middle-aged to young population is small, there is excess demand for consumption by a relatively large cohort of retirees. Retirees do not want to hold financial assets: in fact, they are selling them to fund their consumption. For markets to clear, asset prices have to fall.[15] Abel (2001) uses this intuition to predict that as baby boomers retire, stock prices will decline. The predictions are not, however, clear cut: Brooks (2002), for example, argues that the baby boom effect on asset prices is weak. The second mechanism where demography can predict stock returns is that, since different cohorts have different risk characteristics, asset prices change as the aggregate risk characteristics of the economy change. In an influential study, Bakshi and Chen (1994) show that risk aversion increases as people age and, as the average age rises in the population, the equity premium should increase.

In testing a link between demographic risk and asset returns, it is important to use international data; using only one country's demographic experience is highly suspect because demographic changes are so gradual. The literature employing cross-country analysis includes Erb, Harvey, and Viskanta (1997), Ang and Maddaloni (2005), and Arnott and Chave, (2011). There is compelling international empirical evidence that demography does affect risk premiums.

Political Risk

The last macro risk that an asset owner should consider is *political* or *sovereign risk*. Political risk has been always important in emerging markets: the greater the political risk, the higher the risk premiums required to compensate investors for bearing it. Political risk was thought to be of concern only in emerging markets.[16] The financial crisis changed this, and going forward political risk will also be important in developed countries.[17]

[14] Available at http://www.frbsf.orglcsip/tfp.php.

[15] This is the main economic mechanism in, for example. Geanakoplos, Magill, and Quinzii (2004).

[16] Harvey (2004) finds little evidence, for example, that political risk is reflected in developed countries.

[17] For a recent paper showing how political risk affects equity risk premiums, see Pastor and Veronesi (2012).

2.4 DYNAMIC FACTORS

The CAPM factor is the market portfolio, and, with low-cost index funds, exchange-traded funds, and stock futures, the market factor is tradeable. Other factors are tradeable too. These factors reflect macro risk and at some level should reflect the underlying fundamental risks of the economy. Macro factors like inflation and economic growth, however, are not usually directly traded (at least not in scale, with the exception of volatility), and so dynamic factors have a big advantage that they can be easily implemented in investors' portfolios.

I present examples of dynamic factors using the best-known example of a tradeable multifactor model introduced by Fama and French (1993). I interchangeably use the words "style factors," "investment factors," and "dynamic factors." Sometimes these are also called "smart beta" or "alternative beta," mostly by practitioners.

Fama-French (1993) Model

The Fama-French (1993) model explains asset returns with three factors. There is the traditional CAPM market factor and there are two additional factors to capture a size effect and a value/growth effect:

$$E(r_i) = r_f + \beta_{i,MKT}E(r_m - r_f) + \beta_{i,SMB}E(SMB) + \beta_{i,HML}E(HML), \tag{2.2}$$

where two new factors, *SMB* and *HML*, appear alongside the regular CAPM market factor.

Let us briefly recap the effect of the CAPM market factor (see Chapter 1). When the market does poorly, stocks that have high exposures to the market factor (stocks with high betas, $\beta_{i,MKT}$) also tend to do badly. That is, high beta stocks tend to tank in parallel when the market tanks. But over the long run, the CAPM predicts that stocks with high betas will have higher average returns than the market portfolio to compensate investors for losses when bad times hit—defined by the CAPM theory as low returns of the market.

Robert Merton (1973), Stephen Ross (1976), and others developed the theoretical multifactor model framework in the 1970s, but it took another two decades for an explosion of studies to demonstrate that factors other than the market mattered empirically. Two of these effects—size and value—are in the Fama-French model. Fama and French did not discover these effects; they just provided a parsimonious model to capture their effects. Unfortunately for the original authors, most of the credit for these risk factors now gets assigned to Fama and French.

In Equation (2.2), the first factor in addition to the market factor in the Fama-French model is *SMB*, which refers to the differential returns of small stocks minus big stocks (hence *SMB*), where small and big refer simply to the market capitalization of the stocks. (Fama and French were clearly not marketers, so the labels on the factors are a tad banal.) The *SMB* factor was designed to capture the outperformance of small firms relative to large firms.

The other factor in the Fama-French model is the *HML* factor, which stands for the returns of a portfolio of high book-to-market stocks minus a portfolio of low book to market stocks. The book-to-market ratio is book value divided by market capitalization, or the inverse of equity value normalized by book value. In essence, a value strategy consists of buying stocks that have low prices (normalized by book value, sales, earnings, or dividends, etc.) and selling stocks that have high prices (again appropriately normalized). Academics often normalize by book value. Thus, value stocks are stocks with low prices relative to book value. Growth stocks have high prices relative to book value. The value effect refers to the phenomenon that value stocks outperform growth stocks, on average. One can normalize prices by measures other than book value—which practitioners do when they build their (often proprietary) value factors.

Fama and French's *SMB* and *HML* factors are constructed to be *factor mimicking portfolios*. They are constructed to capture size and value premiums, respectively and use the (CAPM and multifactor) concept of diversification to ensure that the factors capture size and value effects by averaging across many stocks. These factors are long–short portfolios and take positions away from the market portfolio.[18] The average stock, however, only has market exposure since every stock can't be small and every stock can't be large. Let's examine this point more closely because it is intimately related with the profound CAPM insight that the average investor holds the market.

Suppose ~~Buffett~~, I mean Huffet, is a value stock, headed by a manager who likes to buy companies trading for less than their fundamental value, measured, say, by book value. In the Fama-French model in Equation (2.2), Huffet has a positive *HML* beta, $\beta_{i,HML}$. Value stocks, on average, do better than growth stocks. Relative to the CAPM, the expected return on Huffet is adjusted upward by $\beta_{i,HML} \times E(HML)$. Since *HML* is constructed to have a positive risk premium (remember, it goes long high book-to-market stocks, which are value stocks with high returns, and goes short low book-to-market stocks, which are growth stocks with low returns), the Fama-French nudges Huffet's risk premium upward to account for its "valueness."

[18] The *SMB* and *HML* factors are sometimes given as examples of alternative (or smart) beta. I prefer to use the term dynamic factors because technically beta has the strict meaning of measuring exposure to a factor, rather than the factor itself. We invest in factor portfolios, not betas.

Now consider a growth firm, ~~Enron~~, sorry Inron, which has grown rapidly through a series of acquisitions. Inron has a negative *HML* beta. Relative to the CAPM, the expected return on Inron is adjusted downward. since now $\beta_{i,HML} \times E(HML)$ is negative; because Inron is anti-value, or a growth stock, it carries a lower return.

In the Fama-French model, the *SMB* and *HML* betas are centered around zero. The market is actually size neutral and value neutral. Just as the average investor holds the market, the average stock does not have any size or value tilt. It just has market exposure. Furthermore, in the CAPM, the average beta of a stock is one, which is also the beta of the market. The market itself could be affected by macro factors, like GDP growth. inflation, and the factors discussed in the previous sections. The Fama-French model (2.2) prices value stocks like Huffet and growth stocks like Inron relative to the market.

One important assumption in the CAPM and Fama-French models is that the betas are constant. Empirical evidence shows that exposures of some assets to systematic factors vary over time and, in particular, increase during bad times.[19] The variation of betas themselves can be a source of risk. That betas tend to increase during bad times undoubtedly caused the negative returns of risky assets to be larger during the financial crisis than they otherwise would have been had their betas remained constant.

Size Factor

The size effect was discovered by Banz (1981), with similar results in Reinganum (1981), and refers to the fact that small stocks tended to do better than large stocks, after adjusting for their betas. The past tense is appropriate here, because since the mid-1980s there has not been any significant size effect.

Figure 2.5 plots the value of $1 invested in the *SMB* strategy after taking out the market effect beginning in January 1965 to December 2011 in the solid line.[20] The compound returns of *SMB* reach a maximum right around the early 1980s—just after the early Banz and Reinganum studies were published. Since the mid-1980s there has been no premium for small stocks, adjusted for market exposure. International evidence since the mid-1980s has also been fairly weak. Examining international data, Dimson, Marsh, and Staunton (2011) state that if researchers today were uncovering the size effect, "the magnitude of the premium would not command particular attention, and would certainly

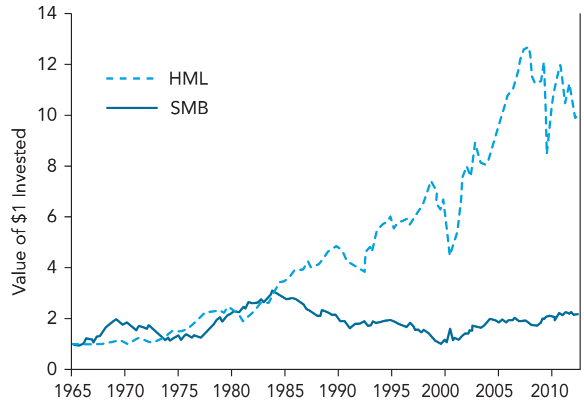


Figure 2.5 Returns to the market-adjusted SMB and HML strategies.

not suggest there was a major 'free lunch' from investing in small caps." Fama and French (2012) also find no size premiums in a comprehensive international data set over recent periods.

There are two responses to the disappearance of the size effect. First, the original discovery of the size premium could have just been data mining. Fischer Black (1993) made this comment immediately after Fama and French's paper was released. The "discovery" of the size effect was then an example of Rosenthal's (1979) "file drawer problem"; which is now a "hard drive problem" (and turning into a "cloud problem"). Researchers store on their hard drives 95% of the results that are statistically insignificant (using a standard *p*-value of 0.05 to judge significance) and only publish the 5% that are statistically significant. The discoverers of the size premium accidentally fell into the 5% category and were just lucky. One telling outcome of data mining is that an effect appears significant *in sample*, where the models are originally estimated, but it fails *out of sample*, where the models are tested after their discovery. Banz's size effect, therefore, might never have truly existed in the first place, and its finding by Banz and Reinganum was pure luck.[21]

The second response is that actually the size effect was there and actions of rational, active investors, acting on news of the finding, bid up the price of small cap stocks until the effect was removed.[22] In this context, the disappearance of the size effect represents the best of the Grossman-Stiglitz (1980) near-efficient market in which practitioners quickly exploit any anomaly. Viewed this way, size does not deserve to be a systematic factor and should be removed from the Fama-French model.

[19] See Ang and Chen (2002).

[20] *SMB*, *HML*, and *WML* data for Figures 2.5 and 2.6 are from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[21] Harvey, Liu, and Zhu (2013) examine hundreds of actors explaining stock returns and investigate the effects of data mining on identifying factors.

[22] Schwert (2003), among others, presents this argument.

It should be noted that small stocks do have higher returns, on average, than large stocks. The effects of other factors, like value and momentum, which we discuss below, are also stronger in small stocks.[23] Small stocks also tend to be more illiquid than large stocks. The pure size effect refers to the possible excess returns of small stocks after adjusting for CAPM betas. The weak size effect today means that an asset owner should not tilt toward small stocks solely for higher risk-adjusted returns. There may only be a case for preferring small caps based on wanting to pursue higher returns without being able to short the market (to remove small caps' market exposures) or an investor could tilt to small caps because she wishes high returns but cannot lever. The unconstrained, investment-only reason for small caps, however, is not compelling.

2.5 VALUE FACTOR

Unlike size, the value premium is robust. Figure 2.5 graphs cumulated returns on the value strategy, *HML*. Value has produced gains for the last fifty years.[24] There are several notable periods where value has lost money in Figure 2.5, some extending over several years: the recession during the early 1990s, the roaring Internet bull market of the late 1990s, and there were large losses from value strategies in the financial crisis over 2007–2008. The risk of the value strategy is that although value outperforms over the long run, value stocks can underperform growth stocks during certain periods. It is in this sense that value is risky.

The benefits of value have been known since the 1930s. Graham and Dodd published a famous book, *Security Analysis*, in 1934, that serves as a guide to identifying firms with low prices relative to their fundamental value. Academics and practitioners proxy fundamental value today by various balance sheet variables or transformations thereof. Graham and Dodd were at Columbia Business School, where I teach, and a strong value-investing tradition continues at my institution today with the Heilbrunn Center for Graham & Dodd Investing. Modern academic research into the value effect began with Basu (1977), and the last few decades have seen an explosion of papers offering various explanations for the value premium. These explanations, like most of the finance literature, fall largely into two camps: the rational and the behavioral.

Rational Theories of the Value Premium

In the rational story of value, value stocks move together with other value stocks after controlling for market exposure (and in fact covary negatively with growth stocks). All value stocks, therefore, tend to do well together or they do badly together. Finding a value stock that doesn't move together with the pack is like finding the sourpuss not thrashing at a heavy metal concert. Just as the euphoria can't last, and some fans experience throbbing migraines the next day, value doesn't always earn high returns. Value is risky, and the riskiness is shared to a greater or lesser degree by all value stocks. Some value risk can be diversified by creating portfolios of stocks, but a large amount of value movements cannot be diversified away. (In fact, Fama and French exploit this *common covariation* in constructing the *HML* factor.) In the context of the APT, since not all risk can be diversified away, the remaining risk must be priced in equilibrium, leading to a value premium.

The Fama-French model itself is silent on why value carries a premium. In contrast, the CAPM provides a theory of how the market factor is priced and even determines the risk premium of the market (see Equation 2.1). To go further, we need to delve into an economic reason for why the value premium exists.

In the pricing kernel formulation, any risk premium exists because it is compensation for losing money during bad times. The key is defining what those bad times are. Let's look at Figure 2.5 again. The bad times for value do not always line up with bad times for the economy. Certainly value did badly during the late 1970s and early 1980s when the economy was in and out of recession. We had a recession in the early 1990s when value also did badly, and the financial crisis in 2008 was unambiguously a bad time when value strategies posted losses. But the bull market of the late 1990s? The economy was booming, yet value stocks got killed. Rational stories of value have to specify their own definitions of bad times when value underperforms, so that value earns a premium on average.

Some factors to explain the value premium include investment growth, labor income risk, nondurable or "luxury" consumption, and housing risk. A special type of "long-run" consumption risk also has had some success in explaining the value premium.[25] During some of the bad times defined by these factors, the

[23] See Loughran (1997) for size-value interactions and Chen, Hong, and Stein (2007) for size-momentum interactions.

[24] Interestingly, a few authors including Ang and Chen (2007) show that value did not exist in the first half of the twentieth century.

[25] For a labor income risk explanation, see Santos and Veronesi (2006). Parker and Julliard (2005) and Lustig and van Nieuwerburgh (2005) consider luxury consumption and housing risk, respectively. For "long-run" consumption risk, see for example, Bansal, Dittmar, and Lundblad (2005).

betas of value stocks increase. This causes value firms to be particularly risky.[26]

Firm Investment Risk

A key insight into the behavior of value and growth firms was made by Berk, Green, and Naik (1999). They build on a *real option* literature where a manager's role is to optimally exercise real investment options to increase firm value.[27] A firm in this context consists of assets in place plus a set of investment options that managers can choose (or not) to exercise. The CAPM is a linear model, and it turns out that CAPM does not fully work when there are option features (see also Chapter 3). Berk, Green, and Naik show that managers optimally exercise investment options when market returns are low. These investment options are dynamically linked to book-to-market (and size) characteristics, giving rise to a value premium. Value firms are risky; their risk turns out to be the same conventional bad times risk as the CAPM or other macro-based factors. Be a value investor only if you can stomach losses on these firms during these bad times.

Lu Zhang has written a series of papers explaining the value premium in terms of how value firms are risky as a result of their underlying production technologies. An important paper is Zhang (2005), which builds on the *production-based asset pricing* framework introduced by Cochrane (1991, 1996). Cochrane teaches us to look at firm investment to study firm returns. The gist of the Cochrane–Zhang story is as follows. Value firms and growth firms differ in how flexible they are and how quickly they can respond to shocks. During bad times, value firms are risky because they are burdened with more unproductive capital. Think of value firms as making stodgy widgets, and when a bad time comes, they cannot shift their firm activities to more profitable activities—they are stuck making widgets. They wish to cut back on capital, but they cannot sell their specialized widget-manufacturing equipment. In economists' jargon, they have *high and asymmetric adjustment costs*. Growth firms, however, can easily divest because they employ hotshot young employees and the great bulk of their capital is human capital, not stodgy widget-making factories. Thus, value firms are fundamentally riskier than growth firms and command a long-run premium.

Rational Implications for Asset Owners

The literature is still debating whether the bad times defined by these theories are truly bad times. But this academic debate is not that relevant to an asset owner; bickering over a Cochrane-Zhang story versus a story about another risk factor is not what the asset owner contemplating a value tilt should be doing.

Remember that the average investor holds the market portfolio. The asset owner should take these rational theories and ask, given that each factor defines a different set of bad times, are these actually bad times for me? If I do not need to eat less during periods when investment growth is low, then this is not as bad a time for me as it is for the average investor. Thus, I have a comparative advantage in holding value stocks and can harvest the value premium. Other investors are not comfortable holding value stocks (and should hold growth stocks instead) because they cannot afford to shoulder the losses generated by value stocks during bad times. Overall, the average investor holds the market even though some investors prefer value stocks and some investors prefer growth stocks. Which type you are—value or growth—depends on your own behavior during each of these bad times.

Behavioral Theories of the Value Premium

Most behavioral theories of the value premium center around investor *overreaction* or *overextrapolation* of recent news. The standard story was first developed by Lakonishok, Shleifer, and Vishny (1994). Investors tend to overextrapolate past growth rates into the future. The posterchild growth stock example at the time of writing is Apple Inc. (AAPL), which has achieved tremendous growth over the last few years by introducing a series of must-have products. Investors mistake Apple's past high growth for future high growth. Growth firms, in general, have had high growth rates. The prices of these firms are bid up too high, reflecting excessive optimism. When this growth does not materialize, prices fall, leading to returns on growth stocks being low relative to value firms. The story here is that value stocks are not fundamentally riskier than growth firms, as in the rational stories. Value stocks are cheap because investors underestimate their growth prospects. Conversely growth firms are expensive because investors overestimate their growth prospects.

The value effect can also be produced by investors with other psychological biases. Barberis and Huang (2001) generate a value effect by employing two psychological biases: *loss aversion* and *mental accounting*. Since investors suffer from losses more than they rejoice from equivalent gains, a loss following a loss is more painful than just a single loss. As for mental accounting, here agents look at each stock individually rather than considering overall gains and losses on their portfolios. The Barberis-Huang story of value is that a high book-to-market ratio stock is one that has achieved its relatively low price as the

[26] For time-varying betas of value stocks, see Lettau and Ludvigson (2001b). Petkova and Zhang (2005), Lewellen and Nagel (2006), Ang and Chen (2007), and Ang and Kristensen (2012).

[27] This literature was started by McDonald and Siegel (1985).

result of some dismal prior performance. This burned the investor who now views it as riskier and thus requires higher average returns to hold the stock.

The crucial question that behavioral models raise is: why don't more investors buy value stocks and, in doing so, push up their prices and remove the value premium, just as investors appear to have done with the size premium (at least according to the Grossman-Stiglitz interpretation)? Put another way, why aren't there more value investors? It can't be ignorance; the message of Graham and Dodd has spread far and wide since the 1930s, as demonstrated by the cult-like fervor of Berkshire Hathaway annual meetings, where value guru Warren Buffett holds court, or at the Graham & Dodd Breakfast conferences at Columbia Business School.

Perhaps investors think value investing is too difficult. Yet simple strategies of academics sorting stocks on a book-to-market basis are available even to the smallest retail investor using stock screens freely available on the Internet. Perhaps it is the legacy of the efficient market theory developed in the 1970s (see Chapter 1) but active managers have never believed in truly efficient markets, and now academics no longer believe in them either.

Maybe not enough institutions have sufficiently long horizons to effectively practice value investing. The value effect documented here, though, is different from the "deep value" practiced by some investors, including Buffett. That requires five- to ten-year horizons. The book-to-market value effect described here is a three- to six-month effect. But perhaps even this horizon is too long for most "long-horizon" investors.

Behavioral Implications for Asset Owners

The relevant question that an asset owner should ask from a behavioral standpoint is simple: do you act like the market, or do you have the ability not to overextrapolate or overreact? If you know you overextrapolate, do you overextrapolate less than the average investor? If you act like everyone else, then simply hold the market portfolio. If you overreact more, perhaps unconsciously, then you tilt toward growth stocks. If you can go against the crowd, then value investing is for you.

Value in Other Asset Classes

Value in essence buys assets with high yields (or low prices) and sells assets with low yields (or high prices). While in equities the strategy is called value-growth investing, the same strategy of buying high-yielding assets and selling low-yielding assets works in all asset classes but goes by different names. Many commentators view these different asset-class strategies as distinct, but they share many features. In fixed income, the value strategy is called *riding the yield curve* and is a form of the duration premium. In commodities it is called the *roll return*, and the sign of the return is related to whether the futures curve is upward- or downward-sloping.

In foreign exchange, the value strategy is called *carry*. This is a popular strategy that goes long currencies with high interest rates and shorts currencies with low interest rates. Traditionally, the former have been countries like Australia and New Zealand, and the latter have been countries like Japan and more recently the United States. In these cases, we can use versions of Equation (2.2) within each asset class. For example, adapting Lustig, Roussanov, and Verdelhan (2011), we could capture the carry (or "value") returns of a foreign currency by using

$$E(FX_i) = \beta_{i,FX} E(HML_{FX}), \tag{2.3}$$

where FX_i is the foreign carry return of country i,$\beta_{i,FX}$ is the loading of currency i on the carry factor HML_{FX}, which is formed by going long currencies with high interest rates minus currencies with low interest rates. There is no conceptual difference between the value strategy in currencies in Equation (2.3) and the value strategy in equities in Equation (2.2) in terms of viewing low prices as equivalent to high yields.

Value strategies turn out to have some common components across asset classes, as shown by Koijen et al. (2012) and Asness, Moskowitz, and Pedersen (2013). While we have compelling stories, both rational and behavioral, of value strategies within equities, bond, and currency markets, we have few theories to link the risk premiums of value strategies across markets.[28] Nevertheless, value is a pervasive factor and theoretically can be implemented cheaply and in size by a large investor. For small investors, there are low-cost index products for value strategies in equity, fixed income, and currency markets as well. The pervasiveness of value across many different asset classes turns out to be something an asset owner should exploit in factor investing.

Momentum

Another standard investment factor is momentum. This burst onto the academic scene with Jegadeesh and Titman (1993) in the same year that Fama and French were capturing size and value factors.[29] Industry professionals like Richard Driehaus, a star mutual fund manager, had already been practicing

[28] Burnside et. al. (2010) develop a disaster-based explanation of the carry trade, similar to the disaster explanations of the equity premium.

[29] Momentum had appeared in the literature with Levy (1967) but was ignored until Jegadeesh and Titman's (1993) work.

momentum for decades.[30] Jegadeesh and Titman (1993) noted that Value Line, a vendor of financial data, has been providing price momentum signals in its publications since the 1980s.

Momentum Investing

Momentum is the strategy of buying stocks that have gone up over the past six (or so) months (winners) and shorting stocks with the lowest returns over the same period (losers), The momentum effect refers to the phenomenon that winner stocks continue to win and losers continue to lose. We call the momentum factor *WML*, for past winners minus past losers. (It is also called *UMD*, for stocks that have gone up minus stocks that have gone down.) The momentum strategy, like size and value, is a cross-sectional strategy, meaning that it compares one group of stocks (winners) against another group of stocks (losers) in the cross section, rather than looking at a single stock over time. Winners and losers are always relative—stocks win or lose relative to each other, and the market as a whole can go up or down.

Momentum returns blow size and value out of the water. Figure 2.6, which plots cumulated returns from January 1965 to December 2011, for *SMB, HML*, and *WML* speaks for itself. The cumulated profits on momentum strategies have been an order of magnitude larger than cumulated profits on size and value. Momentum is also observed in every asset class: we observe it in international equities, commodities, government bonds, corporate bonds, industries and sectors, and real estate.[31] In commodities, momentum is synonymous with commodities trading advisory funds. Momentum is also called "trend" investing, as in "the trend is your friend."

Momentum returns are not the opposite of value returns: in Figure 2.6, the correlation of *HML* with *WML* is only –16%. But many investors who claim that they are growth investors are actually momentum investors, especially mutual funds, as pure growth underperforms value in the long run. There is one sense in which momentum is the opposite of value. Value is a *negative feedback strategy*, where stocks with declining prices eventually fall far enough that they become value stocks. Then value investors buy them when they have fallen enough to have attractive

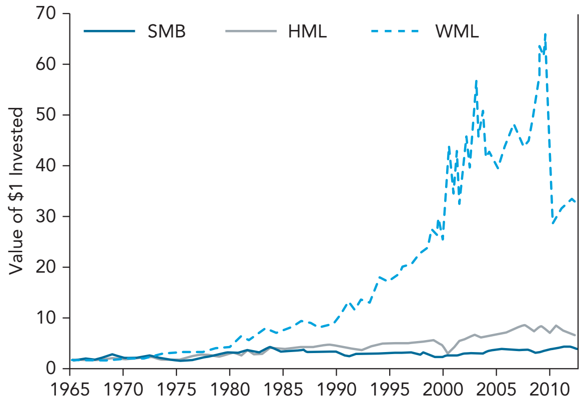


Figure 2.6 **Returns to the SMB, HML and WML strategies.**

high expected returns. Value investing is inherently stabilizing. Momentum is a *positive feedback strategy*. Stocks with high past returns are attractive, momentum investors continue buying them, and they continue to go up! Positive feedback strategies are ultimately destabilizing and are thus subject to periodic crashes, as Figure 2.6 shows and as I discuss below.

Momentum is primarily a *cross-sectional strategy* within an asset class: it looks at a particular group of stocks (those with past high returns) relative to another group of stocks (those with past low returns). Rebalancing, in contrast, should be done primarily at the asset class or strategy level because rebalancing requires the assets or strategies to exist over the long run while individual equities can disappear. Momentum manifests across asset classes, as does value.[32] It can be part of a long-run investor's opportunistic strategy (the Merton (1969) long-run hedging demand portfolio).

Momentum is often used as an investment factor, added onto the Fama-French model:[33]

$$E(r_i) = r_f + \beta_{i,MKT}E(r_m - r_f) + \beta_{i,SMB}E(SMB) + \beta_{i,HML}E(HML) + \beta_{i,WML}E(WML) \qquad (2.4)$$

The same intuition applies as with the Fama-French model. The momentum beta, $\beta_{i,WML}$, is centered around zero. Winner stocks have positive momentum betas; their risk premiums are adjusted upward using Equation (2.4). Loser stocks have negative momentum betas; their risk premiums are adjusted downward. The market, neither a relative winner nor a relative loser, is simply the market.

[30] As recounted by Schwager (1992).

[31] See Asness, Moskowitz, and Pedersen (2012) for momentum in equities, government bonds, currencies, and commodities. The standard momentum effect based on past returns is weak in Japanese equities, but versions of momentum do work in Japan; see Chaves (2012). For momentum in corporate bonds and real estate, see Jostova et al. (2013) and Marcato and Key (2005), respectively. Menkhoff et al. (2012b) is a detailed look at momentum in currencies.

[32] See Blitz and Van Vliet (2008).

[33] Carhart (1997) was the first to do this.

Characterizing Momentum Risk

Figure 2.6 shows that despite the large return, on average, of momentum strategies, momentum is prone to periodic crashes. Some of these have lasted for extended periods. Daniel and Moskowitz (2012) examine these in detail. Of the eleven largest momentum crashes, seven occurred during the Great Depression in the 1930s, one occurred in 2001, and the other three occurred during the financial crisis in 2008. The loser stocks then were tanking financials: Citi, Bank of America, Goldman, and Morgan Stanley, and some others hard hit by circumstances, like General Motors. Loser stocks have a tendency to keep losing, and lose they would have, were it not for Uncle Sam riding to their rescue. Government bailouts put a floor underneath the prices of these stocks, and they consequently skyrocketed. Since momentum strategies were short these stocks, momentum investors experienced large losses when these stocks rebounded. It is notable that the other big momentum drawdowns were concentrated during the Great Depression when policy makers also had great influence on asset prices. Momentum seems to reflect monetary policy and government risk during extraordinary times. These have also been times of high volatility.

What else explains momentum? Tantalizing suggestions in the literature suggest that at least some portion of momentum profits correlates with macro factors. Momentum profits, for example, vary over the business cycle and depend on the state of the stock market, and there is a link with liquidity.[34] In the rational story of momentum (which is still far from being fully fleshed out in the literature), asset owners should examine how they behave facing the various sources of macro risk discussed earlier.

The most widely cited theories are behavioral. In the main behavioral theories, momentum arises because of the biased way that investors interpret or act on information. Suppose good news on a stock comes out. Momentum can be generated in two ways. First, investors could have delayed overreaction to this news, causing the price to persistently drift upward. Second, investors could underreact to the news. The price initially goes up, but it does not go up as much as it should have to fully reflect how good the news actually was. Investors then learn and cause the stock to go up again the next period. Behavioral explanations, then, fall into two camps: momentum is an overreaction phenomenon, or it is an underreaction phenomenon. Distinguishing between these camps is difficult and still bedevils the literature.[35]

The seminal overreaction models are Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998). Barberis, Shleifer, and Vishny's investors suffer from *conservatism bias*, which causes them to overreact to information because they stick doggedly to their prior beliefs. This causes momentum. In the Daniel, Hirshleifer, and Subrahmanyam model, investors also have psychological biases giving rise to momentum. In this model, investors are *overconfident* and overestimate their abilities to forecast firms' future cash flows. They also have biased *self-attribution:* when they are successful, it must be due to their skill, and when they are unsuccessful, it must be due to bad luck. Informed, overconfident investors (think of retail investors and overconfident hedge fund managers) observe positive signals about some stocks that perform well. These overconfident investors attribute the good performance to their own skill, leading to overconfidence. Based on increased confidence, they overreact and push up the prices of stocks above their fundamental values, generating momentum.

The standard reference for the underreaction theory is Hong and Stein (2000). Hong and Stein rely on "bounded rational" investors who have limited information. Momentum in Hong and Stein's model is caused by "news watchers" who receive signals of firm value but ignore information in the history of prices. Other investors trade only on past price signals and ignore fundamental information. The information received by the news watchers is received with delay and is only partially incorporated into prices when first revealed to the market. This causes underreaction.

In both the underreaction and overreaction models, prices eventually reverse when they revert to fundamentals in the long run.

Implications for Asset Owners

In the context of these behavioral stories, the asset owner should think about what types of psychological biases that she has and how these biases differ from those of the average investor. Do you overreact (or underreact) in a way similar to the market? You should also think about how the market's psychological biases can change. Momentum strategies are negatively skewed; the skewness of the momentum strategy in Figure 2.6 is −1.43. At a minimum, the investor should be able to tolerate large drawdowns induced by momentum strategies. Historically, these declines are concentrated in periods when policymakers have interrupted natural progressions of momentum, as in the Great Depression and the financial crisis.

2.6 VALUE INVESTING REDUX

Factor risks represent bad times for an investor. There are two main types of factors—macro factors and investment factors. Assets are exposed to factor risks. The higher the exposure for a

[34] See Chordia and Shivakumar (2002). Cooper, Gutierrez, and Hameed (2004), and Pástor and Stambaugh (2003), respectively.

[35] See, for example, Jegadeesh and Titman (2001).

factor with a positive risk premium (the higher the asset's beta), the higher the asset's expected return.

The value strategy is an example of an investment style factor. In a rational story, value produces losses during bad times, and value stocks are risky. These bad times could coincide with bad times of the economy, as proxied by poor economic growth or poor returns of the market, or they could correspond with bad outcomes of other factors like firm investment. The average investor dislikes these bad times and requires a risk premium to hold value stocks. Thus, value stocks earn high returns to compensate investors for lousy returns during bad times. In behavioral stories, value stocks have high returns because investors underestimate the growth rates of value stocks. They overextrapolate the past growth rates of growth, or glamour, stocks, leading to growth stocks being overpriced and value stocks underpriced. If these behavioral biases are not arbitraged away, value stocks have high excess returns.

3

Alpha (and the Low-Risk Anomaly)

Learning Objectives

After completing this reading you should be able to:

- Describe and evaluate the low-risk anomaly of asset returns.
- Define and calculate alpha, tracking error, the information ratio and the Sharpe ratio.
- Explain the impact of benchmark choice on alpha and describe characteristics of an effective benchmark to measure alpha.
- Describe Grinold's fundamental law of active management, including its assumptions and limitations, and calculate the information ratio using this law.
- Apply a factor regression to construct a benchmark with multiple factors, measure a portfolio's sensitivity to those factors, and measure alpha against that benchmark.
- Explain how to use style analysis to handle time-varying factor exposures.
- Describe issues that arise when measuring alphas for nonlinear strategies.
- Compare the volatility anomaly and the beta anomaly and analyze evidence of each anomaly.
- Describe potential explanations for the risk anomaly.

Excerpt is Chapter 10 of Asset Management: A Systematic Approach to Factor Investing, *by Andrew Ang.*

See bibliography on pp. 195–200.

3.1 CHAPTER SUMMARY

Alpha—the average return in excess of a benchmark—tells us more about the set of factors used to construct that benchmark than about the skill involved in beating it. A positive alpha under one set of factors can turn negative using a different set. Whatever the benchmark, alpha is often hard to detect statistically, especially when adjustments for risk vary over time. The risky anomaly—that stocks with low betas and low volatilities have high returns—appears to be a strong source of alpha relative to standard market-weighted benchmarks and value-growth, momentum, and other dynamic factors.

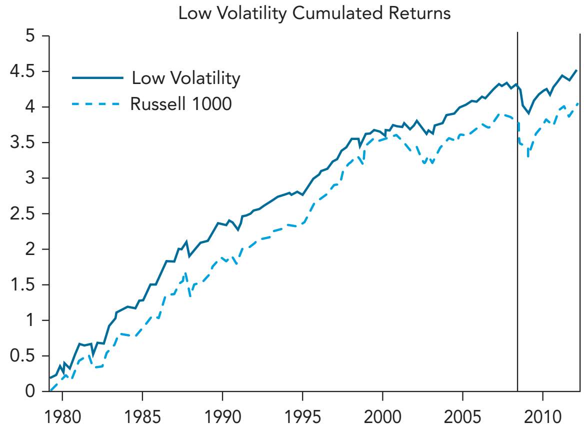


Figure 3.1 Low volatility cumulated return.

3.2 GM ASSET MANAGEMENT AND MARTINGALE

Jim Scott and Brian Herscovici of General Motors (GM) Asset Management, which manages GM Pension Fund, were entranced as Bill Jacques, the CIO of Martingale Asset Management, gave a pitch on his firm's low volatility strategy.[1]

Low volatility seemed too good to be true.

Jacques claimed that stocks with low risk, as measured by past volatility or past beta, had higher returns than stocks with high risk. This was contrary to generally accepted financial theory: the capital asset pricing model (CAPM), for example, stated that there should be a positive relation between risk and return. Martingale's low volatility strategy was constructed to exploit this risk anomaly. Figure 3.1 graphs cumulated returns on Martingale's Low Volatility LargeCap+ strategy based on the Russell 1000 from January 1979 to April 2012. Up until December 2007, the returns are simulated by Martingale's analysts. From January 2008, indicated in Figure 3.1 by the vertical line, performance is live.

Over the full sample, Martingale's volatility strategy beat the Russell 1000 by 1.50%. It had done even better after going live, generating an average outperformance of 1.83% since 2008. Using one-month T-bill rates as the risk-free benchmark, the low volatility strategy had a mean excess return of 8.59% and a standard deviation of 12.22%, which gave the volatility strategy a Sharpe ratio of $8.59/12.22 = 0.70$ over the full sample. In comparison, the Sharpe ratio of the Russell 1000 was 0.45. Adjusted for market risk, Martingale's performance was even more impressive; running a CAPM regression using the Russell 1000 as the market portfolio produced an alpha of 3.44% with a beta of 0.73.

Scott and Herscovici, however, were not immediately swayed by these numbers and Jacques's enthusiastic presentation. Was low volatility a good fit for GM? Tracking error was closely monitored at GM Asset Management, and the Martingale product had a historic tracking error of 6.16% versus the Russell 1000. This was relatively high; other U.S. equity strategies that the group typically considered had tracking errors below 6%. Scott, as managing director of GM Asset Management's global public markets business, insisted that the equities group would need to offset that risk or convince the investment committee that a change of benchmark was warranted. However, what ultimately mattered for GM was meeting its pension fund liability. Herscovici, a global equities portfolio manager, believed that "the low volatility strategy has two main factors consistent with GM's direction to reduce surplus volatility: lower risk and a higher correlation to pension liabilities."

Most of all, Scott and Herscovici wondered what explained this anomaly. If it offered such a great opportunity, why weren't a lot of other investors doing it? And would it persist going forward?

3.3 ACTIVE MANAGEMENT

Alpha is often interpreted as a measure of skill. That is debatable. It is first and foremost a statement about a benchmark.

[1] This is based on the case "GM Asset Management and Martingale's Low Volatility Strategy," 2012, Columbia CaseWorks ID #110315. I have known Scott and Jacques for some time. I have taught a class with Scott on quantitative investments, which is basically what quant hedge funds do. Jacques contacted me out of the blue after I wrote a report on the Norwegian sovereign wealth fund in 2009 with William Goetzmann and Stephen Schaefer. In the report, I cited a Harvard Business School case study about Martingale written by Luis Viceira (another adviser to Martingale). Jacques phoned me one morning, swung by my office later the same day, and then immediately invited me to be part of Martingale's academic advisory board.

Definition of Alpha

Alpha is the average return in excess of a benchmark. Thus, the concept of alpha requires first defining a benchmark against which alpha can be measured.

We define the excess return, r_t^{ex}, as the return of an asset or strategy in excess of a benchmark as

$$r_t^{ex} = r_t - r_t^{bmk}, \tag{3.1}$$

for r_t the return of an asset or strategy and r_t^{bmk} the benchmark return. Sometimes we refer to the excess returns, r_t^{ex}, as *active returns*. This terminology assumes that the benchmark is *passive* and can be produced without any particular investment knowledge or even human intervention. Common passive benchmarks are market-weighted portfolios, like the S&P 500 or the Russell 1000, which investors can track by buying low-cost index funds.

We compute the alpha by taking the average excess return in Equation (3.1):

$$\alpha = \frac{1}{T}\sum_{t=1}^{T} r_t^{ex}, \tag{3.2}$$

where there are T observations in the sample.

The other two terms that we need to define are *tracking error* and the *information ratio*. Tracking error is the standard deviation of the excess return; it measures how disperse the manager's returns are relative to the benchmark:

$$\text{Tracking Error} = \bar{\sigma} = \text{stdev}(r_t^{ex}). \tag{3.3}$$

Tracking error constraints are imposed to ensure a manager does not stray too far from the benchmark. The larger the tracking error, the more freedom the manager has. If the benchmark is risk adjusted, then academics like to call tracking error *"idiosyncratic volatility"* (see the following two sections).

Finally, the information ratio is the ratio of alpha to tracking error:

$$\text{Information Ration} = IR = \frac{\alpha}{\bar{\sigma}}. \tag{3.4}$$

Alpha by itself could be produced by a manager taking large amounts of risk. The information ratio divides the alpha by the risk taken, so it is *the average excess return per unit of risk*. Information ratios above one are not common—although many hedge funds trying to raise money claim to have them. (Since the financial crisis, information ratios on many funds and strategies have come down substantially.)

A special case of Equation (3.4) is when the benchmark is the risk-free rate, r_t^f. (Note that the risk-free rate is known at the beginning of the period and applies from $t - l$ to t.) Then, the alpha is the average return in excess of the risk-free rate,

$\alpha = \overline{r_t - r_{ft}}$ (where the upper bar denotes the sample average of the excess return), and the information ratio coincides with the Sharpe ratio,

$$\text{Sharpe Ratio} = SR = \frac{\overline{r_t - r_{ft}}}{\sigma},$$

where σ is the volatility of the asset.[2]

Benchmarks Matter

Martingale's low volatility strategy is based on the Russell 1000 universe of large stocks, so naturally Jacques takes the Russell 1000 as his benchmark. This is also a convenient benchmark for the asset owner because large investors can buy Russell 1000 index funds for less than 10 basis points per year. Jacques is running an active strategy with relatively high fees and his volatility strategy must offer compelling returns (higher returns, lower risk, or both) relative to the Russell 1000 to attract investors.

A combination of assets or asset classes can also serve as a benchmark. One of Scott and Herscovici's concerns is that Jacques's low volatility strategy has a high tracking error, of 6.16%, relative to the Russell 1000. The tracking error is high because Martingale's product has a low beta of 0.73 relative to the same index. Beta is measured by regressing excess returns of the fund (using T-bills as the risk-free asset) on excess returns of the Russell 1000 (which is the regression implied by the CAPM; see Chapter 1):

$$r_t - r_t^f = 0.0344 + 0.7272(r_t^{R1000} - r_t^f) + \varepsilon_t,$$

where r_t^{R1000} represents the return of the Russell 1000 and ε_t is the residual of the regression.

The $\alpha = 3.44\%$ per year in this CAPM regression is the average excess return of the low volatility strategy relative to a *market-adjusted portfolio*. Put another way, we can rewrite the CAPM regression using a benchmark portfolio of a risk-free asset and 0.73 of the Russell 1000:

$$r_t = 0.0344 + \underbrace{0.2728 r_t^f + 0.7272 r_t^{R1000}}_{r_t^{bmk}} + \varepsilon_t,$$

That is, suppose the benchmark $r_t^{bmk} = 0.27 r_t^f + 0.73 r_t^{R1000}$ consists of a portfolio holding 27% in the risk-free asset and 73% in the Russell 1000. The low volatility strategy outperforms this benchmark by 3.44% per year. The information

[2] In practice, volatilities of excess returns or raw returns are generally almost the same—unless risk-free rates are very volatile, which happens in some emerging markets.

ratio of the low volatility strategy with this *risk-adjusted benchmark* is a very high 0.78.

If we assume a naive benchmark of just the Russell 1000, we falsely assume that the beta of the low volatility strategy is one (when in fact it is 0.73). With the Russell 1000 benchmark we have

$$r_t = 0.0150 + \underbrace{r_t^{R1000}}_{r_t^{bmk}} + \varepsilon_t,$$

so the alpha is 1.50% per year. The information ratio of the low volatility strategy relative to the naive Russell 1000 benchmark is just 0.24. This is not the correct risk-adjusted benchmark because the beta of the low volatility strategy is not one.

Thus, even with a simple Russell 1000 portfolio, failing to adjust the benchmark for risk can make a huge difference in the alpha!

Ideal Benchmarks

What are the characteristics of a sound benchmark? It should be:

1. Well defined

   Produced by an independent index provider, the Russell 1000 is verifiable and free of ambiguity about its contents. Thus it ably defines the "market portfolio" for GM and Martingale.

2. Tradeable

   Alpha must be measured relative to *tradeable benchmarks*, otherwise the computed alphas do not represent implementable returns on investment strategies.

   So the benchmark should be a realistic, low-cost alternative for the asset owner. The Russell 1000 is a natural passive benchmark for Martingale's low volatility strategy because low-cost mutual fund and ETF versions are available.

3. Replicable

   Both the asset owner and the fund manager should be able to replicate the benchmark. Martingale is certainly able to replicate the returns of the Russell 1000 benchmark because it bases its strategy on the Russell 1000 universe. GM Asset Management, the client, is also able to replicate the Russell 1000, either by internally trading it or by employing a Russell 1000 index fund provider. Thus, both Martingale and GM face a common, low-cost option.

   Certain benchmarks can't be replicated by the asset owner because they are beyond the asset owner's expertise. Such nonreplicable benchmarks are not viable choices and make it difficult or impossible to measure how much value a portfolio manager has added because the benchmark itself cannot be achieved by the asset owner. There are some benchmarks, like absolute return benchmarks, that can't even be replicated by the fund manager. In these cases, the fund manager is disadvantaged because she may not even be able to generate the benchmark in the first place.

4. Adjusted for risk

   Sadly, most benchmarks used in the money management business are not risk adjusted.

   Taking the Russell 1000 as the benchmark assumes the beta of Jacques's strategy is one, but the actual beta of the volatility strategy is 0.73. This risk adjustment makes a big difference in the alpha; with the true beta of 0.73, the alpha of the low volatility strategy is 3.44% per year compared to 1.50% when the beta is one.

   When we compute Martingale's beta, we assume that the investor can construct the benchmark portfolio ($r_t^{bmk} = 0.27r_t^f + 0.73r_t^{R1000}$) of T-bills and the Russell 1000 and rebalance it every month. GM Asset Management can easily do this, but some risk-adjusted benchmarks are beyond the reach of less sophisticated clients. We also estimate the beta using data from the whole sample. That is, the beta of 0.73 is the beta generated by Martingale after the fact—at the beginning of the sample, it was not obvious that Martingale would actually trade to generate a beta of 0.73.

   And even though Martingale endorses the Russell 1000 as its benchmark, beta-adjusted or not, what if this is the wrong adjustment for risk? We know that there are more risk factors than just the equity market (see Chapter 2): dynamic factors like value-growth, momentum, credit, and volatility risk also exist. Perhaps we should adjust for some of these as well? In the next section we extend risk adjustments to multiple sources of risk.

Creating Alpha

A portfolio manager creates alpha relative to a benchmark by making bets that deviate from that benchmark. The more successful these bets are, the higher the alpha.

Grinold's (1989) "fundamental law" of active management makes this intuition formal. It states that the maximum information ratio attainable—since it ignores transactions costs, restrictions on trading, and other real-world considerations—is given by:

$$IR \approx IC \times \sqrt{BR}, \tag{3.5}$$

where *IR* is the information ratio; *IC* is the *information coefficient*, which is the correlation of the manager's forecast with the actual returns (how good the forecasts are); and *BR* is the breadth of the strategy (how many bets are taken). Breadth is the number of securities that can be traded and

how frequently they can be traded. High information ratios are generated by a manager finding opportunities—and many of them—where she can forecast well. Grinold and Kahn (1999) state that "it is important to play often (high breadth, *BR)* and to play well (high *IC*)."

The fundamental law has been quite influential in active quantitative portfolio management because it offers a guideline as to how good asset managers have to be at forecasting, how many bets they have to make, or both to generate alpha. Suppose we require an information ratio of 0.5 and we make bets every quarter. A stock market timer trading just the aggregate market has narrow breadth because there are few assets—sometimes just stocks and bonds—to trade. Making only four bets a year, the stock market timer's bets have to be very accurate. She needs an *IC* of 0.25 to obtain an information ratio of $0.50 = 0.25 \times \sqrt{4}$. The cross-sectional strategies of value, size, and momentum (see Chapter 2) have great breadth because hundreds of stocks can be traded. These strategies can have very low *ICs*, often just 2% to 5%, to be highly profitable. Stock selectors doing value, size, or momentum strategies making four hundred independent bets per year need only *ICs* of 0.025 to generate information ratios of $0.50 = 0.025 \times \sqrt{400}$. In short, you can be very talented at forecasting and make those bets really count, or you can have a very small edge but make a lot of bets. Both lead to high information ratios.

Grinold's fundamental law is derived under mean-variance utility, and so all the shortcomings of mean-variance utility apply.[3] In particular, by using mean-variance utility, the Grinold-Kahn framework ignores downside risk and other higher moment risk while assuming that all information is used optimally. A crucial assumption is that the forecasts are independent of each other. Nevertheless, the fundamental law is useful to frame the active management process. Alpha begins with raw information; we process that information into forecasts and then optimally and efficiently construct portfolios that balance return forecasts against risk. The key to generating alpha is forecasting. That requires superior information or superior ability to process public information.

While the fundamental law was intended originally as a portfolio construction tool, Knut Kjær, the founding CEO of Norges Bank Investment Management, which manages the Norwegian sovereign wealth fund, also used the intuition of the fundamental law as a management style. Since alpha is created by finding many different forecasts that are independent, Kjaer's philosophy was to delegate widely to specialized managers. He tried to ensure that they act as independently as possible. He explained:[4]

> To achieve the greatest possible level of independence for decisions on active investments, these must be delegated to many different groups and individuals. This delegation must be real, without intervention from superiors, provided that the individual employee stays within the agreed structure and risk limits. . . . The sense of ownership which derives from responsibility and an absence of intervention from superiors is an important motivation and driving force for the development of this expertise.

There are two very important limitations of the fundamental law. The first is that *ICs* are assumed to be constant across *BR*. The first manager whom you find may truly have a high *IC*, but the one hundredth manager whom you hire probably does not. As assets under management increase, the ability to generate *ICs* diminishes. Indeed, the empirical evidence on active management, shows *decreasing returns to scale:* as funds get bigger. performance deteriorates. This effect is seen in mutual funds, hedge funds, and private equity.[5] Thus, *ICs* tend to fall as assets under management rise.

Second, it is difficult to have truly independent forecasts in *BR*. Manager decisions tend to be correlated and correlated bets reduce *BR*. An equity manager with overweight positions on 1,000 value stocks offset by underweight positions in 1,000 growth stocks has not placed 1,000 different bets; he's placed just one bet on a value-growth factor. Hiring one hundred different fixed income managers who are all "reaching for yield" by buying illiquid bonds gets you not one hundred different bets but rather a single bet on an illiquidity factor. Correlated factor bets tend to dominate at the overall portfolio level—a reason why top-down factor investing is so important.

Despite its influence in industry, Grinold's fundamental law makes only scant appearance in the academic literature. This is because Grinold's framework is a statistical model devoid of economic content. It says nothing about where positive risk-adjusted opportunities are; it only provides a method of systematically evaluating them. I cover a promising area of alpha in an upcoming section.

[3] The fundamental law is itself an approximation in the mean-variance framework. A better approximation to Equation (3.5) is made by Hallerbach (2011). Extensions to the fundamental law allow time-varying signals for the forecasts (Ye (2008)), correlated forecasts (Buckle (2004)), and incorporate estimation risk (Zhou (2008)).

[4] NBIM Annual Report, 2007, "Ten Years of NBIM."

[5] See Chen et. al. (2004), Fung et. al. (2008), and Kaplan and Schoar (2005), respectively.

3.4 FACTOR BENCHMARKS

Consider the CAPM applied to asset (or strategy or fund) *i*:

$$E(r_i) - r_f = \beta_i(E(r_m) - r_f), \tag{3.6}$$

where $E(r_i)$ is the expected return of asset *i*, r_f is the risk-free rate (U.S. T-bills), β_i is the beta of asset *i*, and $E(r_m)$ is the expected return of the market portfolio. Let's assume the beta is 1.3, $\beta_i = 1.3$, so we can write Equation (3.6) as

$$E(r_i) = r_f + 1.3E(r_m) - 1.3r_f,$$

and, rearranging, we have

$$\underbrace{E(r_i)}_{\$1} = \underbrace{-0.3r_f + 1.3E(r_m)}_{\$1}. \tag{3.7}$$

Applying the CAPM assumes that we can produce the same return as \$1 in asset *i* by holding a short position in T-bills of −\$0.30 and a levered position in the market of \$1.30. Note that we have \$1 on both the left-hand side and right-hand side of Equation (3.7). Thus, the CAPM implies a *replicating portfolio:* a combination of risk-free assets and the market has the same expected return as the asset under the CAPM. Thus, the beta of the asset is actually a *mimicking portfolio weight.*[6]

Factor benchmarks are a combination of investment portfolios, or factors, on the right-hand side that give the same return as the asset on the left-hand side. The factor benchmark describes the systematic components of asset *i*'s return.

The alpha of asset *i* is any expected return generated in excess of the short \$0.30 in T-bills and the long \$1.30 position in the market portfolio:

$$E(r_i) = \alpha_i + \underbrace{[-0.3r_f + 1.3E(r_m)]}_{E(r^{bmk})}, \tag{3.8}$$

where the benchmark consists of the risk-adjusted amount held in equities and the risk-free rate, $r^{bmk} = -0.3r_f + 1.3r_m$. Since the benchmark in this case comes from the CAPM, alpha is the average return in excess of the return predicted by the CAPM.

Factor Regressions

We can estimate the risk-adjusted benchmark, or equivalently the mimicking portfolio, using factor regressions.

CAPM Benchmark

To illustrate, consider the grand master of value investing, Warren Buffett. Taking monthly returns on Berkshire Hathaway from 1990 to May 2012, I run the following CAPM regression:

$$r_{it} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \varepsilon_{it}, \tag{3.9}$$

where ordinary least squares assumes that the residuals ε_{it} are independent of the market factor. Estimating the CAPM regression in Equation (3.9) implicitly yields the CAPM-implied mimicking portfolio weights. The estimates for Berkshire Hathaway are

	Coefficient	T-stat
Alpha	0.72%	2.02
Beta	0.51	6.51
Adj R^2	0.14	

This implies that the CAPM benchmark consists of 0.49 in T-bills and 0.51 in the market portfolio, $r^{bmk} = 0.49r_f + 0.51r_m$, and that \$1 invested in Berkshire Hathaway is equivalent to

$$(1 - 0.51) = \$0.49 \text{ in T-bills}$$
$$+ \$0.51 \text{ in the market portfolio.}$$

Relative to this benchmark portfolio, Buffett is adding

$$+ 0.72\% \text{ (alpha) per month.}$$

This is impressive performance! Buffett is generating an alpha of $0.0072 \times 12 = 8.6\%$ per year with a risk of approximately half that of the market ($\beta = 0.51$). The alpha estimated using a market portfolio benchmark in Equation (3.9) is often called Jensen's alpha, after Michael Jensen's pioneering 1968 study of mutual fund performance. (No, mutual funds generally don't beat the market.)

The alpha is also statistically significant, with a high *t*-statistic above two. The cutoff level of two corresponds to the 95% confidence level, a magic threshold for statisticians. Buffett is special. Most factor regressions do *not* produce significant alpha estimates. The adjusted R^2 of the CAPM regression is 14%, which is also relatively high. For most stocks, CAPM regressions produce R^2s of less than 10%.[7] The high R^2 indicates the fit of the CAPM benchmark is very good compared to the typical fit for a CAPM regression for an individual stock.

[6] Equation (3.7), which can be estimated in a regression, takes the perspective of matching \$1 invested in the asset on the left-hand side with the factor benchmark on the right-hand side. An alternative approach is to scale both the left- and right-hand sides so that they have the same volatility. This would match risk exposures rather than dollars invested.

[7] CAPM regressions tend to produce low R^2s because idiosyncratic risk is large in stock returns. Economically, there are tremendous diversification benefits in creating portfolios.

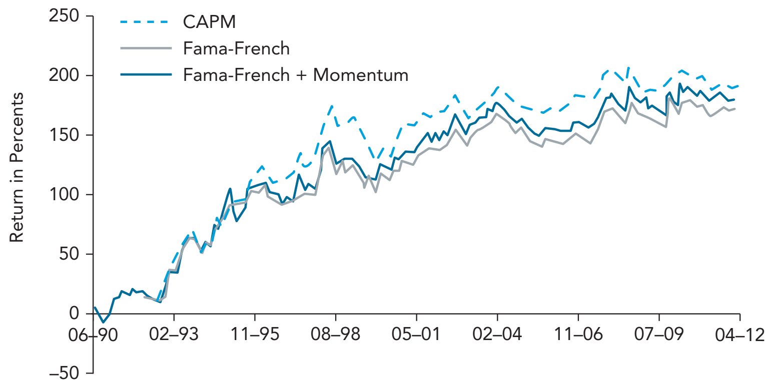


Figure 3.2 **BRK cumulative excess returns.**

The excess return generated by Berkshire Hathaway relative to the CAPM risk-adjusted benchmark is

$$r^{ex} = r_i - r^{bmk}$$
$$= r_i - (0.49r_f + 0.51r_m).$$

The average of the excess returns is the alpha. $\alpha = E(r^{ex}) = 0.72\%$ per month. Figure 3.2 plots the cumulative excess returns of Berkshire Hathaway. The solid line corresponds to the CAPM benchmark. Note that the slope of the cumulative excess returns flattens in the more recent period, especially since the mid-2000s. In 2012, Berkshire Hathaway has a capitalization over \$220 billion. In the early 1990s, Berkshire's market capitalization was less than \$10 billion. As Berkshire Hathaway has grown bigger, its average excess returns have shrunk. Buffett himself said in 2011,

> The bountiful years, we want to emphasize, will never return. The huge sums of capital we currently manage eliminate any chance of exceptional performance. We will strive, however, for better-than-average results and feel it fair for you to hold us to that standard.[8]

Even Berkshire Hathaway is subject to the law of decreasing returns to scale.

Size and Value-Growth Benchmarks

Eugene Fama and Kenneth French, two of the most influential scholars in finance, introduced a benchmark in 1993 that extended the CAPM to include factors that captured a size effect (small companies outperform large companies) and a value-growth effect (value stocks do better than growth stocks). They labeled their size factor *"SMB,"* for small stocks minus big stocks, and their value-growth factor *"HML,"* for high book-to-market stocks minus low book-to-market stocks. Value stocks are stocks with depressed prices, and the *HML* factor normalizes these prices (market value) relative to book value. Hence value stocks are stock with high book-to-market ratios. Chapter 2 covers the Fama-French model and the economics behind the size and value premiums. (Although the size effect is much weaker today, it is still instructive to include *SMB* to explain some of Berkshire Hathaway's performance, as I show below.)

The *SMB* and *HML* factors are long-short factors. They are mimicking portfolios that consist of simultaneous \$1 long and \$1 short positions in different stocks. That is,

$$SMB = \underbrace{\text{\$1 in small caps}}_{\text{Long}} - \underbrace{\text{\$1 in large caps}}_{\text{Short}}$$

and so *SMB* is designed to capture the outperformance of small companies versus large companies. The *HML* factor picks up the outperformance of value stocks versus growth stocks:

$$HML = \underbrace{\text{\$1 in value stocks}}_{\text{Long}} - \underbrace{\text{\$1 in growth stocks}}_{\text{Short}}.$$

The Fama-French benchmark holds positions in the *SMB* and *HML* factor portfolios, along with a position in the market portfolio as in the traditional CAPM.

Fama and French extended the CAPM regression in Equation (3.9) to include a size factor and a value factor. The Fama-French benchmark can be estimated by running the following regression:

$$r_{it} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + sSMB_t + hHML_t + \varepsilon_{it}, \quad (3.10)$$

which adds the *SMB* and *HML* factors to the standard market factor.

The *SMB* and *HML* factor loadings are given by *s* and *h*, respectively. If a stock co-moves neither with small nor large stocks, it's a medium-size stock, and *s* would be zero. As it starts moving with small stocks *s* becomes positive, and if it moves together with large stocks, *s* is negative. Likewise, *h* measures how much a stock is acting like other value stocks: positive *h*s indicate the stock has a value orientation, and negative *h*s indicate the stock is acting like a growth stock. The market itself is neither small nor big and neither value nor growth, so the market has zero *s* and *h* loadings.

[8] Berkshire Hathaway's 2010 shareholder letter.

Estimating the Fama-French regression for Berkshire Hathaway yields the following coefficients:

	Coefficient	T-stat
Alpha	0.65%	1.96
MKT Loading	0.67	8.94
SMB Loading	−0.50	−4.92
HML Loading	0.38	3.52
Adj R^2	0.27	

Buffett's alpha has fallen from 0.72% per month (8.6% per year) with the CAPM benchmark to 0.65% per month (7.8% per year). Controlling for size and value has knocked nearly 1% off Buffett's alpha.

First, note that the market beta has moved from 0.51 in the pure CAPM regression to 0.67 in the Fama-French specification. This is an indication that adding the *SMB* and *HML* factors is doing something—the market beta would stay the same only if the *SMB* and *HML* factors would have *no* ability to explain Buffett's returns.

The *SMB* factor loading in the Fama-French regression is $s = -0.50$. The negative sign indicates that Berkshire Hathaway is acting the opposite way from a small stock (remember, *SMB* is long small stocks and short large stocks). That is, Berkshire Hathaway has large stock exposure. Note that being large counts against Buffett's outstanding performance because large stocks, according to the Fama-French model, tend to underperform small stocks.

The *HML* loading of $h = 0.38$ says that Berkshire Hathaway has a strong value orientation; it tends to move together with other value stocks.

Thus, the negative *SMB* and positive *HML* factor loadings suggest that Berkshire Hathaway is a large, value investor. Duh, of course it is! It doesn't take the finance cognoscenti to know that this is the investing technique that Buffett has touted since founding Berkshire Hathaway in the 1960s. It is comforting that an econometric technique yields the same result as common sense. But the statistical technique gives us the appropriate benchmark to compute Buffett's risk-adjusted alpha.

The surprising result in the Fama-French regression is that Buffett is still generating considerable profits relative to the size- and value-factor controls: Buffett's monthly alpha of 0.65% is still outsized; the Fama-French model reduces the CAPM alpha by less than 1% per year. This is not because the size and value factors are inappropriate risk factors. Quite the contrary. The Fama-French regression has an adjusted R^2 of 27%, which is large by empirical finance standards, and much higher than the adjusted R^2 of 14% in the CAPM benchmark. The size and value factors, therefore, substantially improve the fit relative to the CAPM benchmark. Buffett's performance is clearly not merely from being a value investor, at least the way value is being measured relative to the CAPM.

The benchmark implied by the Fama-French regression estimates is:

$$\begin{aligned}(1 - 0.67) &= \$0.33 \text{ in T-bills}\\ &+ \$0.67 \text{ in the market portfolio}\\ &- \$0.50 \text{ in small caps}\\ &+ \$0.50 \text{ in large caps}\\ &+ \$0.38 \text{ in value stocks}\\ &- \$0.38 \text{ in growth stocks}\end{aligned}$$

In addition to this benchmark, Buffet is generating

$$+\ 0.65\% \text{ (alpha) per month.}$$

Again, the factor loadings can be translated directly to a benchmark portfolio, only now the portfolio contains (complicated) long-short positions in small/large and value/growth stocks. But it still represents $1 of capital allocated between factor portfolios. Every time we run a factor regression, we are assuming that we can create a factor benchmark portfolio.

Adding Momentum

The momentum effect—that stocks with high returns in the past continue their upward trend and stocks with lousy past returns continue to deliver lousy returns—can be added to the factor benchmark. Momentum is observed in many asset classes and is a systematic factor (see Chapter 2). Buffett famously eschews momentum investing, basing his investment decisions on a company's fundamentals instead of past growth and price movements. Three of his famous quotes are:

> The investor of today does not profit from yesterday's growth.
>
> Focus on the underlying conditions that cause price, rather than price itself.
>
> It's far better to buy a wonderful company at a fair price than a fair company at a wonderful price.

We add a momentum factor, *UMD*, constructed by taking positions in stocks that have gone <u>u</u>p <u>m</u>inus stocks that have gone <u>d</u>own, to the Fama-French benchmark:[9]

$$r_{it} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + sSMB_t + hHML_t + uUMD_t + \varepsilon_{it}, \tag{3.11}$$

where the new *UMD* factor has a loading (or beta) of u. Estimating this regression, we have

[9] First done by Carhart (1997).

	Coefficient	T-stat
Alpha	0.68%	2.05
MKT Loading	0.66	8.26
SMB Loading	−0.50	−4.86
HML Loading	0.36	3.33
UMD Loading	−0.04	−0.66
Adj R^2	0.27	

These estimates are very close to the Fama-French regression estimates. Consistent with Buffett's avowed eschewal of momentum investing, the UMD loading is close to zero ($u = -0.04$) and statistically insignificant. Note that the adjusted R^2 of this regression is 27%, exactly the same as the Fama-French regression, implying that adding the momentum factor has not improved the fit of the factor regression. Buffett's alpha has even improved slightly by adding the momentum factor ($\alpha = 0.68\%$ per month) compared to the Fama-French regression ($\alpha = 0.65\%$ per month).

For completeness, the mimicking portfolio implied by this Fama-French plus momentum benchmark is:

$(1 - 0.66) = \$0.37$ in T-bills
$+$ \$0.66 in the market portfolio
$-$ \$0.50 in small caps $+$ \$0.50 in large caps
$+$ \$0.36 in value stocks $-$ \$0.36 in growth stocks
$-$ \$0.04 in past winning stocks $+$ \$0.04 in past losing stocks

Buffett is also adding:

$+0.68\%$ (alpha) per month.

Figure 3.2 shows cumulative excess returns relative to the Fama-French benchmark in the dashed line and the Fama-French plus momentum benchmark in the solid line. Both lie below the CAPM benchmark, which is mainly a consequence of lowering Buffett's alpha by including the *HML* value-growth factor.

Doing without Risk-Free Assets

Benchmark portfolios need not include risk-free assets.

CalPERS

CalPERS is the largest public pension fund in the United States and had \$246 billion of assets at June 30, 2011.[10] A benchmark for this pension fund might be a passive portfolio of index funds in stocks and bonds—the benchmark that the Canada Pension Plan has adopted through its Reference Portfolio. A stock-bond benchmark can be run extremely cheaply—for close to zero—and is a viable yardstick for judging whether active management is adding value.

A benchmark regression for CalPERS' returns would be

$$\underbrace{r_{it}}_{\$1} = \alpha + \underbrace{\beta_s r_{st} + \beta_b r_{bt}}_{\$1} + \varepsilon_{it}, \qquad (3.12)$$

where r_{it} is the return of CalPERS, r_{st} is the S&P 500 equity market return, and r_{bt} is a bond portfolio return—in this case, the Ibbotson Associates long-term corporate bond total return index. To obtain a benchmark portfolio, we require the restriction that

$$\beta_s + \beta_b = 1.$$

That is, the portfolio weights must sum to one. Then, \$1 placed into CalPERS on the left-hand side of Equation (3.12) can be replicated by a portfolio of stocks and bonds (with portfolio weights, which also must sum to one) on the right-hand side, plus any alpha generated by the CalPERS' funds manager.

Estimating Equation (3.12) on CalPERS' annual returns from 1990 to 2011, we get the following:

	Coefficient	T-stat
Alpha	−1.11%	−1.16
Bond Loading	0.32	13.97
Stock Loading	0.68	13.97
Adj R^2	0.90	

The high adjusted R^2 of 90% is amazing: CalPERS' returns are extremely well explained by this mimicking portfolio of 32% bonds and 68% stocks!

The point estimate of CalPERS alpha is negative, at 21.11% per year. Should we immediately fire the CalPERS funds manager and put everything into low-cost index funds? Formally, we can only make the statement that "we fail to reject the hypothesis that CalPERS adds value relative to the 32% bonds/68% stocks benchmark portfolio at the 95% level" because the *t*-statistic is less than two in absolute value.

CalPERS, however, is an expensive fund. In 2011 its internal estimate of its expense ratio was upward of 0.50%, while expense ratios inferred from its annual reports exceed 0.80%. (What a travesty that it does not explicitly report its expense ratio in its annual report!) These expense ratios are much higher than those of industry peers. The median expense ratio

[10] Data and additional information for this section is from "California Dreamin': The Mess at CalPERS," Columbia CaseWorks, #120306 and "Factor Investing: The Reference Portfolio and Canada Pension Plan Investment Board," Columbia CaseWorks #120302.

of the largest pension plans studied by Bauer, Cremers, and Frehen (2009) was 0.29%; at the largest 30% of defined benefit plans, the expense ratios are just 0.15%. Thus, CalPERS is three to four times more expensive than the median fund in Bauer, Cremers, and Frehen's sample, and nearly nine times more expensive than the largest 30% of pension plans! Expense ratios for managing typical index stock or bond funds at CalPERS' scale are way below 0.10%. (The Norwegian sovereign wealth fund had an expense ratio of 0.06% in 2012.) So yes, given that CalPERS could run a benchmark stock-bond portfolio for close to zero, perhaps it should consider firing its funds manager and going completely index.

Figure 3.3 plots cumulative excess returns for CalPERS. The estimated 32% stocks/68% bonds benchmark portfolio is shown in the solid line and a standard 40% stocks/60% bonds portfolio is overlaid for comparison. Note the similarity. CalPERS performance improves during 2000–2007. But during the financial crisis in 2008, things completely fall apart, and the fund's performance continued to deteriorate in 2010 and 2011. A large part of this dismal showing was due to CalPERS' failure to rebalance in 2008 and 2009: it sold equities rather than buying them when prices were low.

Real Estate

Canada Pension Plan considers real estate to have many characteristics in common with fixed income and equities—so much in common that the plan doesn't consider real estate a separate asset class. But can real estate exposure be replicated by a factor portfolio of stocks, bonds, and, potentially, listed REITs, which offer indirect real estate exposure?

Real estate returns are complicated because they are not tradeable. Leaving aside this problem, I take quarterly real estate returns from the National Council of Real Estate Investment Fiduciaries from June 1978 to December 2011 (my left-hand side variable). I consider factor benchmark regressions using S&P 500 stock returns, Ibbotson long-term corporate bond returns, and the FTSE NAREIT index returns (my right-hand side variables).

I run the following factor regressions:

$$\begin{aligned} r_{it} &= \alpha + \beta_{REIT}REIT_t + \beta_b r_{bt} + \varepsilon_{it}, \\ r_{it} &= \alpha + \beta_b r_{bt} + \beta_s r_{st} + \varepsilon_{it}, \\ r_{it} &= \alpha + \beta_{REIT}REIT_t + \beta_b r_{bt} + \beta_s r_{st} + \varepsilon_{it}, \end{aligned} \tag{3.13}$$

where $REIT_t$ is the return to the NAREIT portfolio consisting of traded REITs, r_{bt} is the bond return, and r_{st} is the stock return, which have factor loadings of β_{REIT}, β_b, and β_s, respectively. We require that, in all cases, the factor loadings add up to one so that they can be interpreted as a factor portfolio benchmark.

	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
Alpha	−0.51%	−1.02	−0.43%	−0.90	−1.50%	−1.05
REIT Loading	0.30	5.92			0.12	1.81
Bond Loading	0.70	14.0	0.65	12.7	0.26	3.75
Stock Loading			0.35	6.95	0.61	11.6

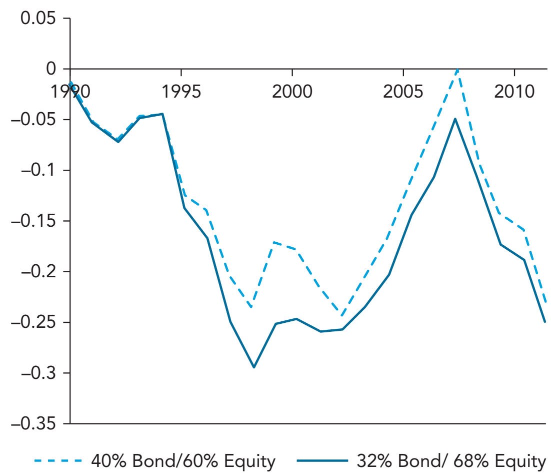


Figure 3.3 **CalPERS cumulated excess returns.**

The estimated coefficients are shown in the table above.

For all of these factor benchmarks, direct real estate does not offer significant returns in excess of a factor benchmark. In fact, the point estimates are negative and around 0.50% per quarter. Interestingly, the factor benchmark consisting of just bonds and stocks indicates that the optimal combination of stocks and bonds to mimic real estate is 35% stocks and 65% bonds.

Figure 3.4 graphs cumulative excess returns of direct real estate relative to these factor benchmarks. While there was some value added in the early 1980s relative to these REIT, bond, and stock factors, the factor benchmarks did much better than direct real estate from the mid-1980s to the early 2000s. Direct real estate picked up relative to the factor benchmarks in the mid-2000s, coinciding with the period's property boom. Figure 3.4 clearly shows the crash in real estate markets in 2008 and 2009 toward the end of the sample.

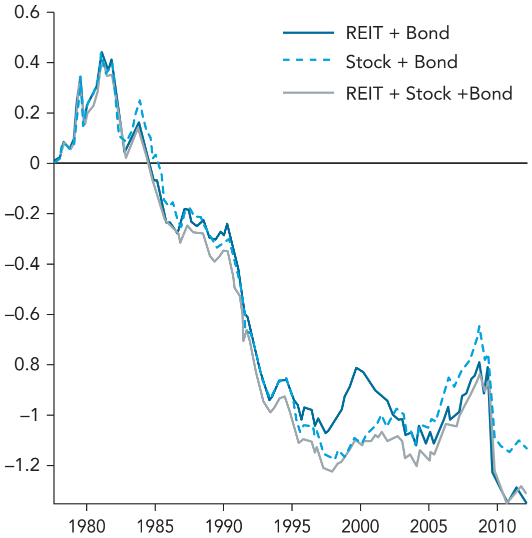


Figure 3.4 Cumulative NCREIF excess returns.

Time-Varying Factor Exposures

William Sharpe, one of the inventors of the CAPM, introduced a powerful framework to handle time-varying benchmarks in 1992. He called it *"style analysis."* In our context, style analysis is a factor benchmark where the factor exposures evolve through time.[11]

To illustrate time-varying factor exposures in the spirit of Sharpe's style analysis, consider four funds:

LSVEX: LSV Value Equity. LSV is a "quantitative value equity manager providing active management for institutional investors" and was named after its founding academics: Josef Lakonishok, Andrei Shleifer, and Robert Vishny.[12]

FMAGX: Fidelity Magellan. One of the most famous retail mutual funds, it grew to prominence under superstar manager Peter Lynch in the 1980s and 1990s.

GSCGX: Goldman Sachs Capital Growth. How can we not include a Goldman Sachs name?

BRK: Berkshire Hathaway. Since we've been using Buffett's example, let's stay with it.

I use monthly data from January 2001 to December 2011.

[11] Computing standard errors for alphas when factor loadings vary over time, and even when the alphas themselves vary over time, is tricky, as Ang and Kristensen (2012) show. For a summary of style analysis, see Horst, Nijman, and de Roon (2004).

[12] See http://www.lsvasset.com/about/about.html.

Here are the Fama-French and momentum factor regressions using *constant* factor weights:

	LSVEX	FMAGX	GSCGX	BRK
Alpha	0.00%	−0.27%	−0.14%	0.22%
t-stat	0.01	−2.23	−1.33	0.57
MKT Loading	0.94	1.12	1.04	0.36
t-stat	36.9	38.6	42.2	3.77
SMB Loading	0.01	−0.07	−0.12	−0.15
t-stat	0.21	−1.44	−3.05	−0.97
HML Loading	0.51	−0.05	−0.17	0.34
t-stat	14.6	−1.36	−4.95	2.57
UMD Loading	0.2	0.02	0.00	−0.06
t-stat	1.07	1.00	−0.17	−0.77

The only alpha that is statistically significant is Fidelity Magellan, which is −0.27% per month or −3.24% per year. Poor investors in Fidelity lose money, and their losses are statistically significant. Berkshire Hathaway's alpha estimate is positive but insignificant. Our analysis in section 3.1 had a significantly positive alpha but we started in 1990. Now, starting ten years later in 2001, we don't even obtain statistical significance for Buffett. Detecting statistical significance of outperformance is hard, even in samples of more than ten years.

Looking at the factor loadings, LSV seems to be a big value shop—with a large *HML* loading of 0.51 (with a massive *t*-statistic of 14.6). Berkshire Hathaway is still value, too, with an *HML* loading of 0.34. Fidelity is a levered play on the market, with a beta of 1.12. Since none of the *UMD* loadings are large or significant, none of these funds are momentum players.

Style analysis seeks to rectify two potential shortcomings of our analysis so far:

1. The Fama-French portfolios are not tradeable.[13]
2. The factor loadings may vary over time.

Style Analysis with No Shorting

Style analysis tries to replicate the fund by investing passively in low-cost index funds. The collection of index funds that replicate the fund is called the *"style weight."*

[13] GM Asset Management has implemented tradeable versions of the Fama-French portfolios. See Scott (2012) for further details. Cremers, Perajisto, and Zitzewitz (2012) argue that the nontradeability of the Fama-French indices leads to distortions in inferring alpha.

To illustrate, let's take the following index ETFs:

SPY: SPDR S&P 500 ETF, which is designed to mimic the S&P 500;

SPYV: SPDR S&P 500 Value ETF, which tracks the S&P 500 value index; and

SPYG: SPDR S&P 500 Growth ETF, which replicates the S&P 500 growth index.

These low-cost index ETFs are tradeable, unlike the Fama-French portfolios. They belong to the SPDR (pronounced "spider") family of ETFs sponsored by State Street Global Advisors.

Our benchmark factor regression for fund *i* (but I avoid the *i* subscripts to make the notation clearer) is

$$r_{t+1} = \alpha_t + \beta_{SPY,t}SPY_{t+1} + \beta_{SPYV,t}SPYV_{t+1} + \beta_{SPYG,t}SPYG_{t+1} + \varepsilon_{t+1}, \qquad (3.14)$$

where we impose the restriction

$$\beta_{SPY,t} + \beta_{SPYV,t} + \beta_{SPYG,t} = 1,$$

so that the factor loadings, or factor weights, sum to one. The factor weights on the right-hand side of Equation (3.14) constitute a replicating portfolio for fund *i*.

The main idea with style analysis is that we use actual tradeable funds in the factor benchmark. I used SPDR ETFs in Equation (3.14), but I could have used other ETFs or index mutual funds for the benchmark portfolio.

Note the timing in Equation (3.14). The weights are estimated using information up to time *t*. The return of the fund over the next period, $t + 1$, is equal to the replicating portfolio formed at the beginning of the period at time *t* plus a fundspecific residual, $\varepsilon_{t+1,}$ and the fund alpha, α_t for that period. The weights can change over time. Equation (3.14) asks, "Can we find a robot that makes time-varying investments in SPY, SPYV, and SPYG that, together, match the returns of Buffett?"

Figure 3.5 graphs the factor weights (or style weights) of the four funds. I estimate the factors using data over the previous sixty months, $t - 60$ to *t*, to form the benchmark weights at time *t*. In addition to imposing that the factor weights sum to one, I also constrain the factor weights to be all positive (so there is no shorting). The first factor weight is estimated at January 2006.

Panel A of Figure 3.5 shows that LSV is merely a combination of the market (SPY) and value (SPYV). Fidelity Magellan starts off in 2006 as a combination of all three ETFs but, at the end of 2012, ends up being all growth (SPYG). Goldman's growth fund is mostly market exposure (SPY) and growth (SPYG) at the beginning of the sample and at the end of the sample is just growth (SPYG). Buffett's factor exposure is the most interesting. He starts off in 2006 being strongly value (SPYV). During the financial crisis, he switches styles to become growth. Then as the crisis subsides, he goes back to being a strong value manager.

The excess return for $t + 1$ is the return of the fund at the end of the period, *t* to $t + 1$, minus the benchmark portfolio formed using the weights at time *t:*

$$r^{ex}_{t+1} = r_{t+1} - \underbrace{[\beta_{SPY,t}\, SPY_{t+1} + \beta_{SPYV,t}\, SPYV_{t+1} + \beta_{SPYG,t}\, SPYG_{t+1}]}_{r_{t+1}^{bmk}}.$$

I graph the excess returns in Panel B of Figure 3.5. The cumulated excess returns are zero for LSV. Fidelity Magellan's returns trend downward (recall that Magellan significantly subtracts value in the full-sample regressions). Goldman's growth fund also has zero cumulative excess returns. The only fund with an upward trend is Berkshire Hathaway.

Style Analysis with Shorting

What if we allow shorting? In Figure 3.6, I allow the investor to take short positions in the ETFs. I use the following factor regression:

$$r_{t+1} - r_{f,t+1} = \alpha_{i,t} + \beta_{SPY,t}(SPY_{t+1} - r_{f,t+1}) + h_t(SPYV_{t+1} - SPYG_{t+1}) + \varepsilon_{t+1}. \qquad (3.15)$$

This is the "ETF version" of the Fama-French (1993) regression that we estimated in Equation (3.10), without the *SMB* factor, except that we allow the factor loadings to change over time. The *SPYV-SPYG* is an investment that goes long the value SPYV ETF and simultaneously shorts the growth SPYG ETF. Thus, it is analogous to the *HML* factor.

The factor loadings plotted in Panel A of Figure 3.6 show the strong value bias of LSV; with a positive *h* loading on the *SPYV-SPYG* factor. Magellan becomes more of a growth fund over time, with increasingly negative *h* loadings, as does Goldman's growth fund. Berkshire Hathaway's changing factor loadings from value to growth to value can be seen in its negative *h* loadings during 2008 and 2009.

Allowing shorting does not much change the cumulated excess returns in Panel B of Figure 3.6. But allowing shorting, not surprisingly, reduces the alphas. Magellan's trend line for cumulated excess returns becomes more negative when shorting is allowed. Although Buffett's excess returns are positive, they are shifted downward in Figure 3.6, Panel B, compared to the corresponding long-only picture in Figure 3.5, Panel B.

My final comment is that the problems of statistical inference with time-varying portfolio benchmarks are serious. It is hard enough to detect statistical significance with constant portfolio

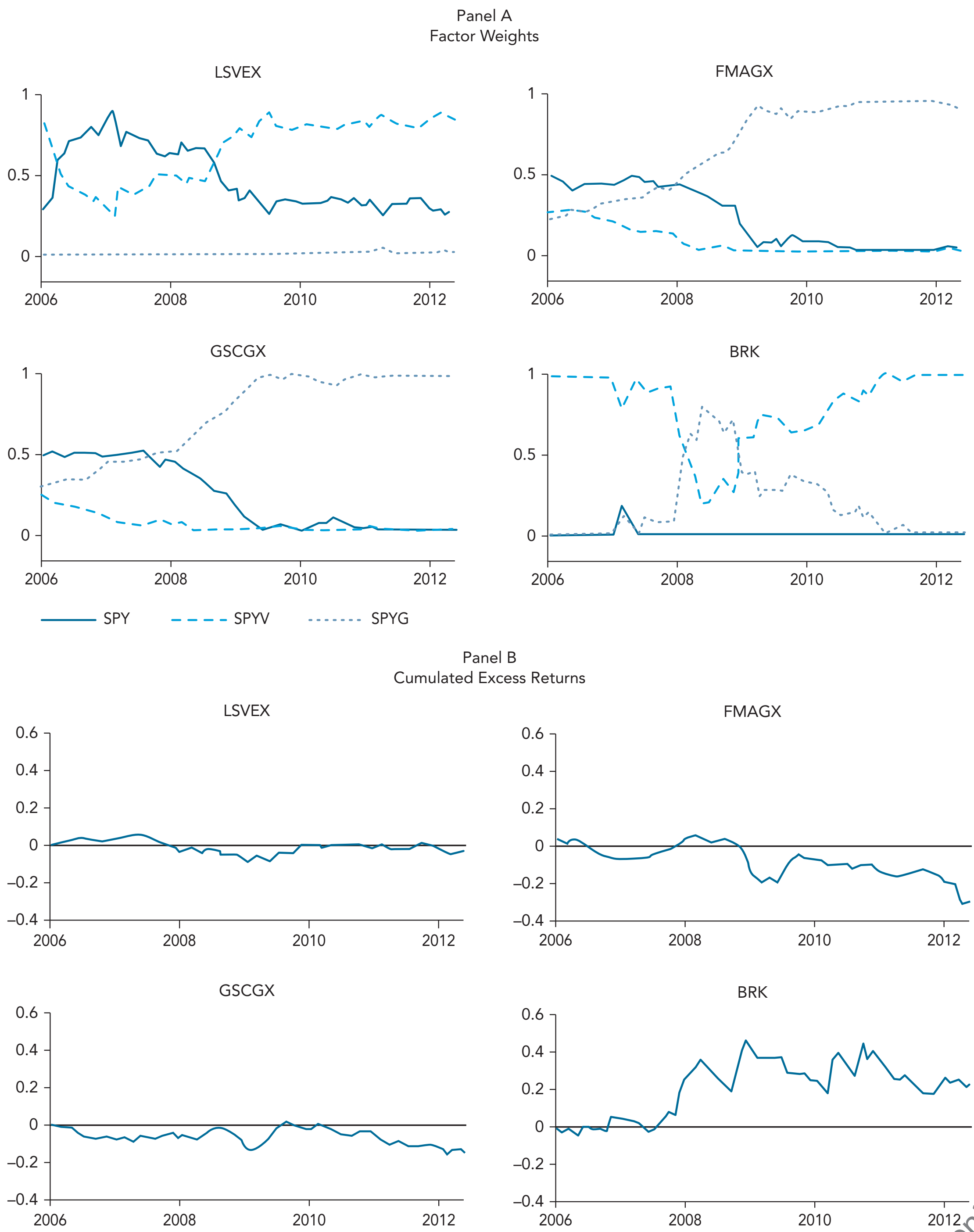

Panel A
Factor Weights
LSVEX
FMAGX
GSCGX
BRK
1
0.5
0
2006
2008
2010
2012
SPY
SPYV
SPYG
Panel B
Cumulated Excess Returns
LSVEX
FMAGX
GSCGX
BRK
0.6
0.4
0.2
0
−0.2
−0.4

Figure 3.5

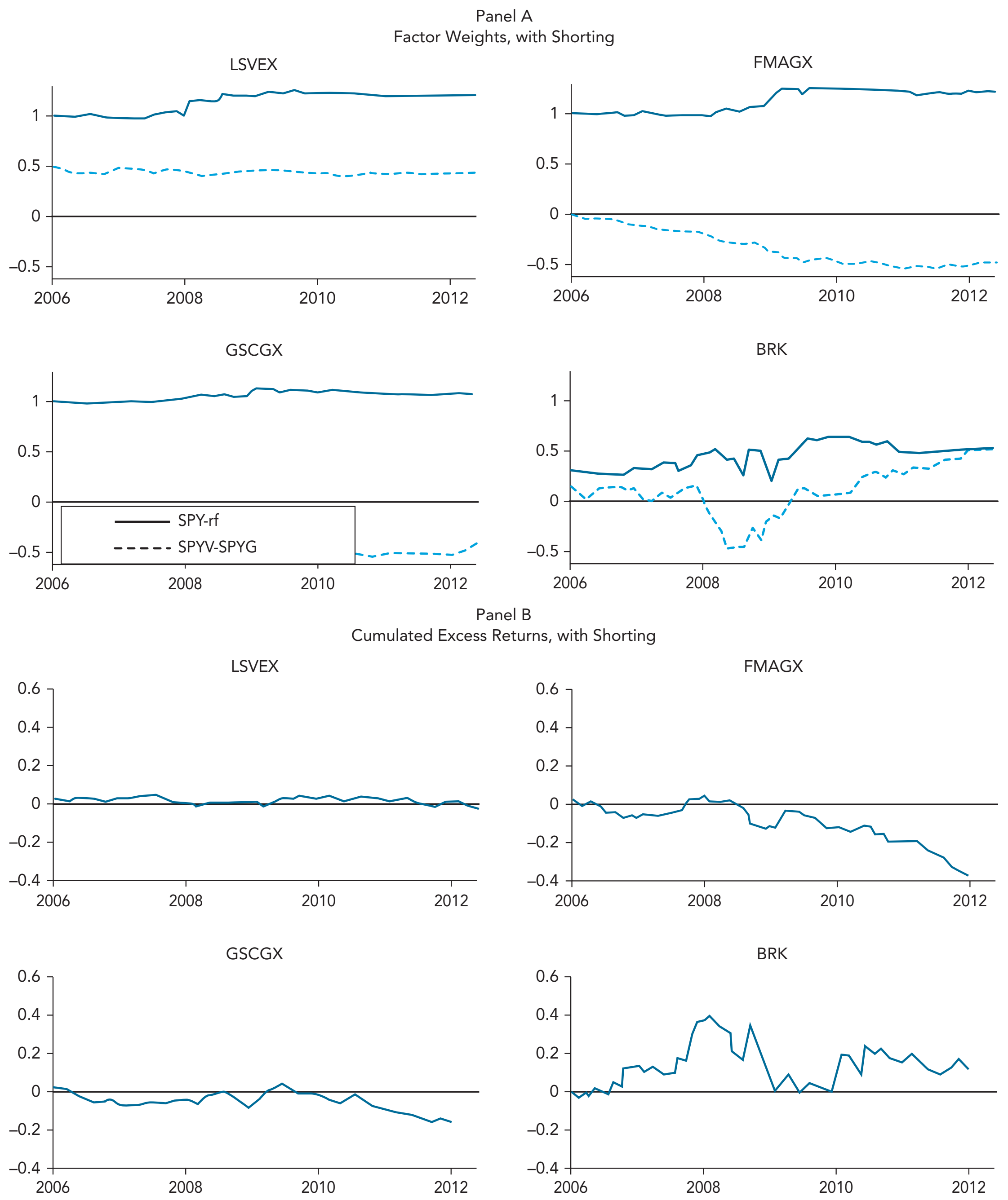

Panel A
Factor Weights, with Shorting
LSVEX
FMAGX
GSCGX
BRK
SPY-rf
SPYV-SPYG
Panel B
Cumulated Excess Returns, with Shorting
LSVEX
FMAGX
GSCGX
BRK
2006
2008
2010
2012

Figure 3.6

benchmarks, and estimated time-varying styles will have even larger standard errors.[14]

Non-Linear Payoffs

With alphas and information ratios, any manager can appear to have talent when he actually doesn't.

Alphas are computed in a *linear* framework. There are many *non-linear* strategies, especially those involving dynamic option strategies, that can masquerade as alpha.[15] To give an extreme (and admittedly stylized) example, consider Figure 3.7. It is produced by selling put options on the market portfolio in a small sample, just using Black-Scholes (1973) prices. The returns on these put options are recorded with crosses. I then run a CAPM regression with these simulated returns. The "alpha" appears to be positive—ta da!—but we know that in the Black-Scholes world there is no extra value created in puts or calls. The alpha is purely illusory. This is not a result of a small sample, even though small samples exacerbate the problem. *No nonlinear strategy can be adequately captured in a linear framework.*[16] This is a serious problem because many common hedge fund strategies, including merger arbitrage, pairs trading, and convertible bond arbitrage, have payoffs that resemble nonlinear option volatility strategies.[17]

Why do dynamic, nonlinear strategies yield false measures of alpha? Because buying and selling options—or any dynamic strategy—changes the distribution of returns.[18] Static measures, like the alpha, information, and Sharpe ratios, capture only certain components of the whole return distribution. Often, short volatility strategies can inflate alphas and information ratios because they increase negative skewness. These strategies increase losses in the left-hand tails and make the middle of the distribution "thicker" and appear to be more attractive to linear performance measures. Skewness and other higher moments are not taken into account by alphas and information ratios.

There are two ways to account for nonlinear payoffs.

[14] See comments by DiBartolomeo and Witkowski (1997).

[15] This was first shown in a seminal paper by Dybvig and Ingersoll (1982). Technically, this is because any factor model used (Dybvig and Ingersoll used the CAPM) implicitly allows the pricing kernel to be negative and permits arbitrage (see Chapter 1). The linear pricing kernel correctly prices all benchmark assets (stocks) but incorrectly prices nonlinear payoffs like derivatives.

[16] For a more formal treatment see Lhabitant (2000), to (2001), and Guasoni, Huberman, and Wang (2011). You can *always* beat a market Sharpe ratio (or information ratio) by selling volatility.

[17] See Mitchell and Pulvino (2001), Gatev, Goetzmann, and Rouwenhorst (2006), and Agarwal et. al. (2011), respectively.

[18] This is true even of simple rebalancing.

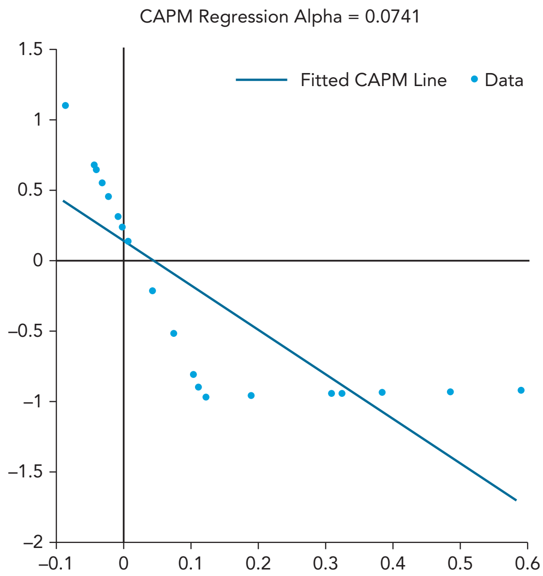


Figure 3.7

Include Tradeable Nonlinear Factors

Aggregate market volatility risk is an important factor, discussed in Chapter 7, and an easy way to include the effects of short volatility strategies is to include volatility risk factors. Other nonlinear factors can also be included in factor benchmarks. By doing so, the asset owner is assuming that she can trade these nonlinear factors by herself. Sometimes, however, the only way to access these factors is through certain fund managers. Controlling for nonlinear factors crucially changes the alphas of hedge funds. Fung and Hsieh (2001), among many others, show that hedge fund returns often load significantly on option strategies.

Examine Nontradeable Nonlinearities

It is easy to test whether fund returns exhibit nonlinear patterns by including nonlinear terms on the right-hand side of factor regressions. Common specifications include quadratic terms, like r_t^2, or option-like terms like $\max(r_t, 0)$.[19] The disadvantage is that, after including these terms, you do not have alpha—we always need tradeable factors on the right-hand side to compute alphas.

But we must move beyond alpha if we want evaluation measures that are robust to dynamic manipulation. These will not be alphas, but they can still be used to rank managers and evaluate

[19] These can be traced to Treynor and Mazuy (1966) and Henriksson and Merton (1981), respectively.

skill. One state-of-the-art measure has been introduced by Goetzmann et. al. (2007).[20] With long enough samples, this measure cannot be manipulated in the sense that selling options will not yield a false measure of performance.

The Goetzmann et. al. evaluation measure is

$$\frac{1}{1-\gamma}\ln\left(\frac{1}{T}\sum_{t=1}^{T}(1+r_t-r_{ft})^{1-\gamma}\right), \tag{3.16}$$

where γ is set to three. Funds can be ranked on this measure from high to low, with the best funds having the highest values. Equation (3.16) is indeed a CRRA or power utility function. (More precisely, it's the certainty equivalent of a CRRA utility function.) Goetzmann et. al. report that Morningstar uses a variant of this measure:

$$\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{(1+r_t-r_{ft})^2}\right)^{-\frac{1}{2}}-1,$$

which is also a CRRA utility function with risk aversion $\gamma = 2$.

Does Alpha Even Exist?

Since alpha is based on a benchmark and estimates of alpha are very sensitive to that benchmark, is there even such a thing as true alpha? It could just be a wrong benchmark. *The* academic literature calls this a *joint hypothesis problem*, and the search for alpha is the same as the testing for market efficiency.[21] In a major contribution, Hansen and Jagannathan (1997) show that it is always possible to find an ex-post benchmark portfolio that produces no alpha. This is less useful ex ante, but it shows that a benchmark portfolio can always be constructed where no alpha exists after the fact. Since Grossman and Stiglitz (1980), the profession recognizes that perfectly efficient markets cannot exist (see Chapter 1)—so there is alpha—but as the analysis of this section has shown, even for a recognized master of investing like Buffett, alpha can be very hard to detect statistically.

The joint hypothesis problem—that alpha and the benchmark are simultaneously determined—is the key problem for asset owners. It is of little use for an academic to say that Fidelity has no alpha, when the asset owner cannot access the complicated size, value-growth, and momentum factors used to compute that alpha. For that asset owner, Fidelity may be providing alpha. For another asset owner, Fidelity may well be adding negative alpha because she can do all the appropriate factor exposure (and implement the underlying replicating factor benchmark portfolios) on her own.

Choosing the right set of factors, then, is the most relevant issue for asset owners. Alpha is primarily a statement about the factor benchmark (or lack of a factor benchmark). We now have enough knowledge of risk adjustments to judge different alpha opportunities, and so we turn to one source of alpha that has recently stirred up debate.

3.5 LOW RISK ANOMALY

The low-risk anomaly is a combination of three effects, with the third a consequence of the first two:[22]

1. Volatility is negatively related to future returns;
2. Realized beta is negatively related to future returns; and
3. Minimum variance portfolios do better than the market.

The risk anomaly is that risk—measured by market beta or volatility—is negatively related to returns. Robin Greenwood, a professor at Harvard Business School and my fellow adviser to Martingale Asset Management, said in 2010, "We keep regurgitating the data to find yet one more variation of the size, value, or momentum anomaly, when the Mother of all inefficiencies may be standing right in front of us—the risk anomaly."

History

The negative relation between risk (at least measured by market beta and volatility) and returns has a long history. The first studies showing a negative relation appeared in the late 1960s and early 1970s.[23] Friend and Blume (1970) examined stock portfolio returns in the period 1960–1968 with CAPM beta and volatility risk measures. They concluded (my italics):

> The results are striking. In all cases risk-adjusted performance is dependent on risk. *The relationship is inverse and highly significant.*

[20] For some other notable manipulation-free performance measures, see Glosten and Jagannathan (1994) and Wang and Zhang (2011).

[21] For a summary of this large literature, see Ang, Goetzmann, and Schaefer (2011).

[22] Some references for the third are Haugen and Baker (1991), Jagannathan and Ma (2003), and Clarke, de Silva, and Thorley (2006). I cover references for the others later.

[23] In addition to the papers in the main text, also see Pratt (1971), Sodolfsky and Miller (1969), and Black (1972).

Haugen and Heins (1975) use data from 1926 to 1971 and also investigate the relation between beta and volatility risk measures and returns. They report (my italics):

> The results of our empirical effort do not support the conventional hypothesis that risk—systematic or otherwise—generates a special reward. Indeed, our results indicate that, over the long run, *stock portfolios with lesser variance in monthly returns have experienced greater average returns than "riskier" counterparts.*

Must of these results were forgotten. But these old results recently have come roaring back.

Volatility Anomaly

I was fortunate to write one paper that helped launch the new "risk anomaly" literature in 2006 with Robert Hodrick, one of my colleagues at Columbia Business School, and two of our former students, Yuhang Xing and Xiaoyan Zhang, who are now professors at Rice University and Purdue University, respectively. We found that the returns of high-volatility stocks were "abysmally low." So low that they had zero average returns. This paper now generates the most cites per year of all my papers and has spawned a follow-up literature attempting to replicate, explain, and refute the results.[24]

First, should there even be a relation between volatility and returns? The whole point of the CAPM and the many multifactor extensions (see Chapter 2) was that stock return volatility itself should not matter. Expected returns, according to these models, are determined by how assets covary with factor risks. Idiosyncratic volatility, or tracking error (see Equation (3.3)), should definitely *not* have any relation to expected returns under the CAPM. But in markets that are segmented due to clientele effects—where some agents cannot diversify or where some agents prefer to hold some assets over others for exogenous reasons—idiosyncratic volatility should be positively related to returns. Intuitively, agents have to be paid for bearing idiosyncratic, risk, resulting in a positive relation between idiosyncratic risk and volatility in equilibrium. In later models with "noise traders," who trade for random reasons unrelated to fundamental valuation, higher volatilities are associated with higher risk premiums.[25]

The Ang et. al. (2006) results show exactly the opposite.

> Particularly notable is the robustness of the negative relation between both idiosyncratic and total volatility with returns. We employed a large number of controls for size, value, leverage, liquidity risk, volume, turnover, bid-ask spreads, co-skewness risk, dispersion in analysts' forecasts, and momentum. We also did not find that aggregate volatility risk explained our result—even though volatility risk is a pervasive risk factor (see Chapter 2). In subsequent work, Ang et. al. (2009), we showed that the volatility effect existed in each G7 country and across all developed stock markets. We also controlled for private information, transactions costs, analyst coverage, institutional ownership, and delay measures, which recorded how fast information is impounded into stock prices. Skewness did not explain the puzzle.

Lagged Volatility and Future Returns

To see the volatility anomaly, I take U.S. stocks, rebalance quarterly from September 1963 to December 2011, and form quintile portfolios. I construct monthly frequency returns. I sort on idiosyncratic volatility using the Fama-French (1993) factors with daily data over the past quarter. (Ranking on total volatility produces very similar results.) I market weight within each quintile similar to Ang et. al. (2006, 2009).

In Figure 3.8, I report the mean and standard deviations of the quintile portfolios on the left-hand axis in the two bars. The volatilities increase going from the low- to high-volatility quintiles, by construction. The average returns are above 10% for the first three quintiles, fall to 6.8% for quintile 4, and then plummet to 0.1% for the highest volatility stocks. High volatility stocks certainly do have "abysmally low" returns. The right-hand axis reports raw Sharpe ratios, which are the ratios of the means to the standard deviations. These monotonically decline from 0.8 to 0.0 going from the low- to high-volatility quintiles.

Contemporaneous Volatility and Returns

Do stocks with high volatilities also have high returns over the same period used to measure those volatilities?

I examine this question in Figure 3.9 by forming portfolios at the end of the period based on realized idiosyncratic volatilities.

[24] Volatility makes many appearances, of course, in tests of cross-sectional asset pricing models before Ang et. al. (2006), but most of them are negative results or show a slight positive relation. For example, in Fama and MacBeth's (1973) seminal test of the CAPM, volatility is included and carries an insignificant coefficient. Eric Falkenstein (2012) recounts that he uncovered a negative relation between volatility and stock returns in his PhD dissertation in 1994, which was never published.

[25] For clientele models, see Merton (1981). For noise trader models, see Delong et. al. (1990).

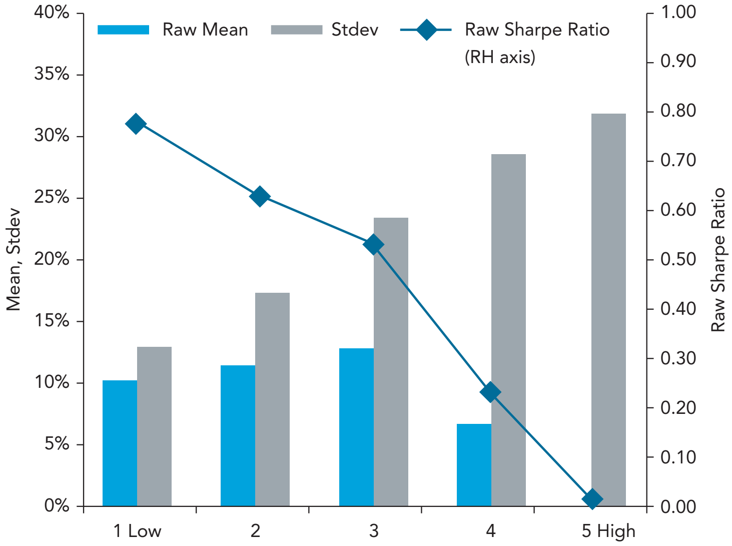


Figure 3.8 Volatility portfolios.

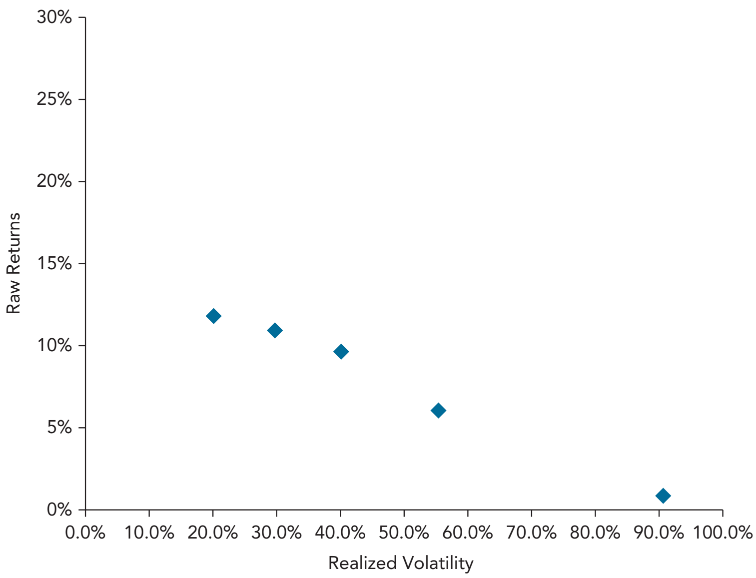


Figure 3.9 Contemporaneous volatility portfolios.

Beta Anomaly

The first tests of the CAPM done in the 1970s did find positive relations between beta and expected returns, but they did not find that pure forms of the CAPM worked. Black, Jensen, and Scholes (1972), for example, found the relation between beta and returns to be "too flat" compared with what the CAPM predicted, but at least the relation was positive.

Fama and French wrote a major paper in 1992 that struck at the heart of the CAPM. While their main results showed that size and value effects dominated beta in individual stocks, they noted that "beta shows no power to explain average returns." In fact, their estimated relation between beta and returns was statistically insignificant. Worse, the point estimates indicated that the relation between beta and returns was negative.

Lagged Beta and Future Returns

In Figure 3.10, I form quintile portfolios rebalancing every quarter based on betas estimated over the previous quarter using daily returns. The portfolios are equal weighted so as to form the largest differences in returns and Sharpe ratios, and returns are at the monthly frequency.

The beta anomaly is that stocks with high betas tend to have lower risk-adjusted returns. Panel A of Figure 3.10 shows that the average returns across the beta quintiles are approximately flat, around 15% for the first four quintiles and slightly lower at 12.7% for quintile 5. The beta anomaly is not that stocks with high betas have low returns—they don't. Stocks with high betas have high volatilities. This causes the Sharpe ratios of high beta stocks to be lower than the Sharpe ratios of low-beta stocks. The right-hand axis of Panel A shows that the raw Sharpe ratios drop from 0.9 to 0.4 moving from the low- to the high-beta quintile portfolios.

I then measure realized returns over the same period. Note that these are not tradeable portfolios. Figure 3.9 plots the average realized volatility and average realized returns of these quintile portfolios; there is still a negative relation between contemporaneous volatility and returns. Thus, the most volatile stocks *currently* lose money (which we cannot forecast), and they also tend to lose money in *the future* as well (which is predictable).

In Panel B of Figure 3.10, I plot the pre-ranking and post-ranking betas. The pre-ranking beta is the beta over the previous three months, which is used to rank the stocks into portfolios. The post-ranking beta is the realized beta over the next three months after the portfolios have been formed. Panel B graphs the average pre-ranking betas of each portfolio with the average post-ranking betas. There is considerable

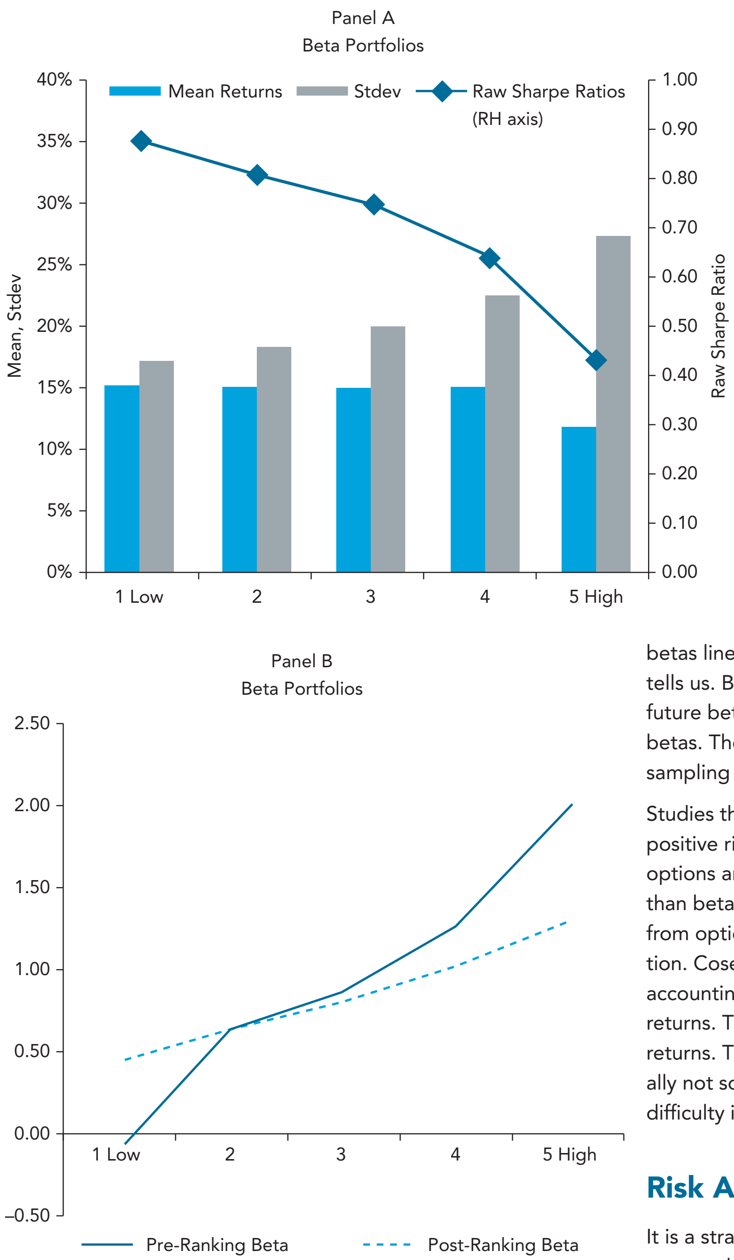


Figure 3.10

noise in estimating betas at both ends, which is why the post-ranking beta line is much flatter than the pre-ranking betas. Betas are noisy! There is, however, still a large spread in post-ranking betas of over 1.0 between the highest and lowest beta portfolios.

Contemporaneous Beta and Returns

The CAPM does *not* predict that lagged betas should lead to higher returns. The CAPM actually states that there should be a *contemporaneous* relation between beta and expected returns. That is, stocks with higher betas should have higher average returns over the same periods used to measure the betas and the returns (see Chapter 2 for more on factor theory).

Figure 3.11 examines the contemporaneous relation between betas and average returns by graphing average realized returns and average realized betas of portfolios formed at the end of each three-month period. It shows, perhaps surprisingly, that there is a positive contemporaneous relation between beta and returns.[26] This is exactly what the CAPM predicts!

Can we reconcile the negative relation between past betas and future returns and the positive contemporaneous relation between betas and realized returns? If we could find the future beta, future betas line up with future returns in keeping with what the CAPM tells us. But Figure 3.10, Panel B, shows that it is hard to predict future betas. Past betas do a lousy job at predicting future betas. There is large variation in betas, and there is substantial sampling error.[27]

Studies that estimate betas from other information tend to find positive risk relations. Buss and Vilkov (2012) estimate betas from options and find them to be better predictors of future betas than betas estimated from past returns. Their betas estimated from option-implied information yield a positive risk-return relation. Cosemans et. al. (2012) use valuation information from accounting balance sheets to compute betas along with past returns. They also estimate a positive relation between betas and returns. Thus, the real mystery in the low-beta anomaly is actually not so much that beta does not work; it is that we have such difficulty in predicting future betas, especially with past betas.

Risk Anomaly Factors

It is a straightforward extension to use these portfolio results to create a benchmark factor for the risk anomaly.

[26] See also Ang, Chen, and Xing (2006). Consistent with the early studies like Black, Jensen, and Scholes (1972), Figure 3.11 also shows that the estimated security market line is "too flat," especially near the y-axis.

[27] Blume (1975) was one of the first to document this. For formal statistics for calculating the paths of time-varying alphas and betas and then standard errors, see Ang and Kristensen (2012).

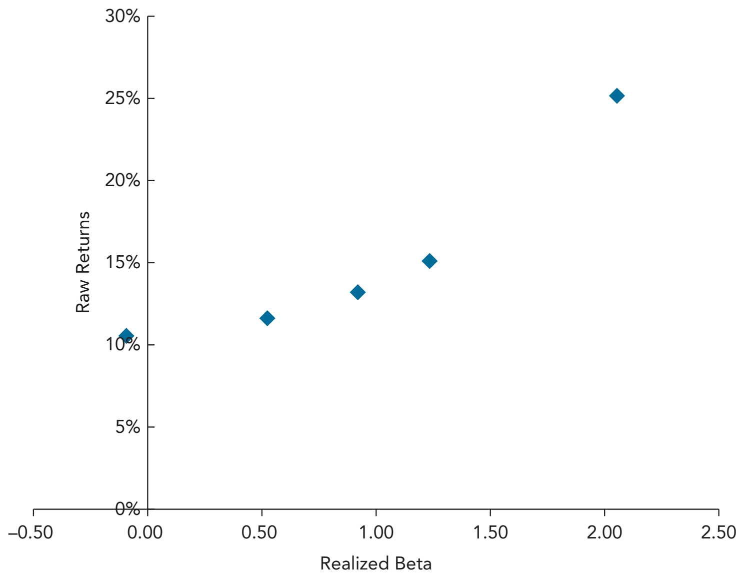


Figure 3.11 **Contemporaneous beta portfolios.**

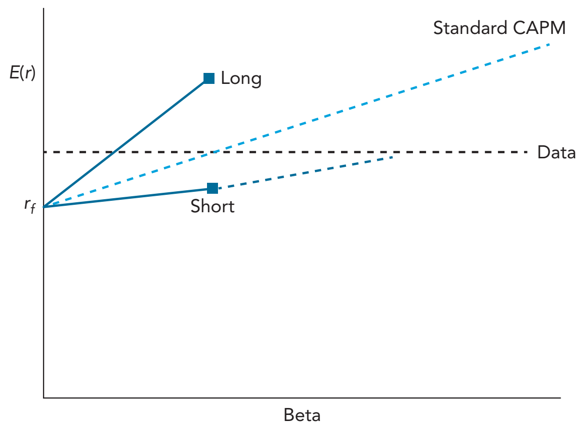


Figure 3.12

Betting against Beta

Frazzini and Pedersen (2010) construct a betting-against-beta *(BAB)* factor that goes long low-beta stocks and short high-beta stocks. Constructing a factor to trade the beta anomaly cannot be done just by taking differences of the portfolios in Figure 3.10. Remember, the differences in average returns across the beta quintiles are tiny—what's large are the differences in Sharpe ratios across betas. Frazzini and Pedersen form their *BAB* factor by scaling the low- and high-beta portfolios by their betas:

$$BAB_{t+1} = \frac{r_{L,t+1} - r_f}{\beta_{L,t}} - \frac{r_{H,t+1} - r_f}{\beta_{H,t}}, \qquad (3.17)$$

where $r_{L,t+1}$ is the return of the low-beta portfolio and $r_{H,t+1}$ is the return of the high-beta portfolio. The betas of the low-beta and high-beta portfolio at the beginning of the period (the pre-ranking betas) are given by $\beta_{L,t}$ and $\beta_{H,t}$, respectively.

Figure 3.12 shows what is going on. The horizontal line labeled "Data" is the empirical pattern of flat average returns with lagged betas in contrast to the upward-sloping line that is predicted by the "Standard CAPM." The long position in the low-beta portfolio is levered. It takes the position where it hits the "Data" line and levers it up to the square marked "Long." The short position in the high-beta portfolio is marked "Short." The *BAB* portfolio does not take the entire position in the high-beta portfolio; it combines it with the risk-free asset to obtain the position marked Short. In effect, the Long and Short portfolios are unit beta positions in the low- and high-beta portfolios.

Frazzini and Pedersen use just two beta portfolios in creating their *BAB* factor. They have little choice. In Figure 3.10, the first quintile portfolio has a pre-ranking beta close to zero. Levering up this portfolio results in a position close to infinity. Thus, one is forced to create very small numbers of portfolios—two or three at most—in the *BAB* factor. One advantage of the volatility portfolios is that they can be directly traded without using the risk-free asset because there are pronounced differences in expected returns, not only volatilities, across the volatility quintiles.

Volatility Factor

I construct a volatility factor, *VOL*, similar to Frazzini and Pedersen's *BAB:*

$$VOL_{t+1} = \sigma_{\text{target}} \times \left(\frac{r_{L,t+1} - r_f}{\sigma_{L,t}} - \frac{r_{H,t+1} - r_f}{\sigma_{H,t}} \right), \qquad (3.18)$$

where $\sigma_{L,t}$ and $\sigma_{H,t}$ are the pre-ranking volatilities of the low- and high-volatility portfolios. While the *BAB* factor scales to unit betas, the *VOL* factor scales to a target volatility. I use the first and fifth quintile portfolios with returns $r_{L,t}$ and $r_{H,t}$, respectively. I set the target volatility $\sigma_{\text{target}} = 15\%$.

Betting-against-Beta and Volatility Factors

Figure 3.13 compares the *BAB* and *VOL* factors from October 1963 to December 2011.[28] The cumulative returns of the *VOL*

[28] I construct a *BAB* factor similar to Frazzini and Pedersen (2012) except I do not follow their step in shrinking the betas. Specifically, betas are computed in one-year rolling regressions using daily frequency returns. There are two beta portfolios created at the end of each month, and the *BAB* factor is constructed using Equation (3.14) using the pre-ranking portfolio betas.

factor are higher than *BAB*, and the volatility factor has a slightly higher Sharpe ratio (0.6 vs. 0.5), but the two factors are largely comparable. The main surprising result is that the beta and volatility effects are very lowly correlated; the correlation between *BAB* and *VOL* is −9%. The volatility and beta anomalies, therefore, are distinct.

Running a Fama-French plus momentum factor regression, we obtain

	BAB Factor		VOL Factor	
	Coeff	T-stat	Coeff	T-stat
Alpha	0.33%	1.89	0.42%	4.37
MKT Loading	−0.17	−4.13	0.87	38.8
SMB Loading	0.29	5.20	−0.63	−20.3
HML Loading	0.48	7.85	0.20	5.73
UMD Loading	0.09	2.35	0.13	6.00

The alpha for the *BAB* factor is 0.33% per month (4% per year) and the *t*-statistic of 1.89 corresponds to a *p*-value of 0.06. So this is borderline statistically significant at the standard 95% level. The *VOL* factor's alpha is slightly higher, at 0.42% per month (5% per year) but is much more statistically significant with a *t*-statistic of 4.37. Note that both *BAB* and *VOL* have significant value tilts (positive *HML* loadings) and momentum tilts (positive *UMD* loadings). The big difference is that the *BAB* factor carries a negative *SMB* loading, whereas it is positive for the *VOL* factor. That is, the beta anomaly manifests more in small stocks. In contrast, the volatility anomaly is more pervasive in large stocks, which are usually easier to trade because they are more liquid.

So should you do low volatility, or should you do low beta? This is not an either-or choice. You should do both.

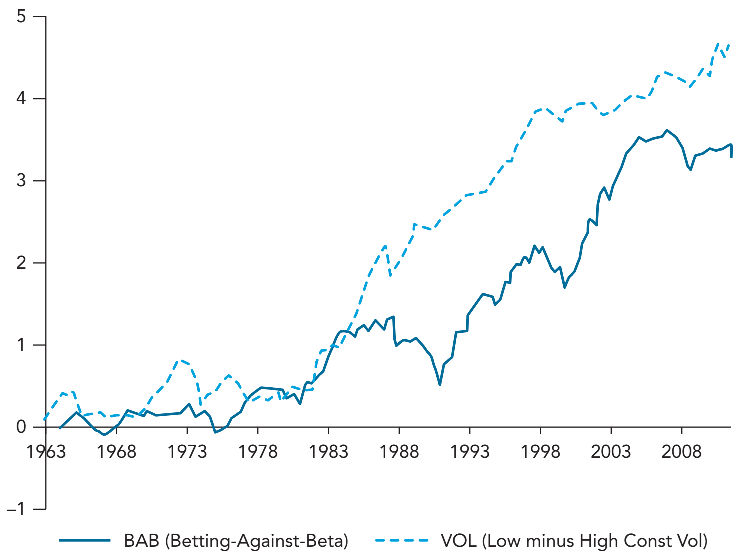


Figure 3.13 Beta and volatility cumulated returns.

Explanations

We are still searching for a comprehensive explanation for the risk anomaly. In my opinion, the true explanation is a combination of all of the explanations follow, plus potentially others being developed.

Data Mining

Some papers in the literature rightfully point out some data mining concerns with the original results in Ang et. al. (2006). There is some sensitivity in the results to different portfolio weighting schemes and illiquidity effects.[29] For the most part, however, the low-risk anomaly is fairly robust. A recent survey article by Chen et. al. (2012) argues that "idiosyncratic volatility is a common stock phenomenon" and is not due to microstructure or liquidity biases.

The best argument against data mining is that the low-risk effect is seen in many other contexts. Ang et. al. (2006) show that the effect appears during recessions and expansions and during stable and volatile periods. Ang et. al. (2009) show that it takes place in international stock markets. Frazzini and Pedersen (2011) show that low-beta portfolios have high Sharpe ratios in U.S. stocks, international stocks, Treasury bonds, corporate bonds cut by different maturities and credit ratings, credit derivative markets, commodities, and foreign exchange. Cao and Han (2013) and Blitz and de Groot (2013) show that the low-risk phenomenon even shows up in option and commodity markets, respectively. Low risk is pervasive.

[29] See Bali and Cakici (2008) and Han and Lesmond (2011), respectively.

Leverage Constraints

Many investors are leverage constrained—they wish to take on more risk but are unable to take on more leverage.[30] Since they cannot borrow, they do the next best thing—they hold stocks with "built-in" leverage, like high-beta stocks. Investors bid up the price of high-beta stocks until the shares are overpriced and deliver low returns—exactly what we see in data. In CAPM parlance, the voracious demand of leverage-constrained investors for high-beta stocks flattens the security market line (see Chapter 1). The leverage constraint story, however, does not explain the underpricing of low-beta stocks relative to the market, only the overpricing of high-beta stocks. Thus, it cannot explain why low-beta or low-volatility assets have higher returns than the market portfolio, but it can explain why some low-beta assets have positive alphas. This story also predicts that leverage-constrained institutions should be attracted to high-risk stocks. In reality, though, institutional investors tend to underweight high-risk stocks; stocks with high idiosyncratic volatility are predominantly held and traded by retail investors.[31]

Agency Problems

Many institutional managers can't or won't play the risk anomaly. In particular, the use of market-weighted benchmarks itself may lead to the low volatility anomaly.[32]

Figure 3.14 draws a theoretical relation between beta and expected returns in the diagonal solid line marked "CAPM" (the security market line). The data relation between returns and beta is the horizontal line marked "Data." Consider stock A, which has positive alpha, and B, which has negative alpha. Unconstrained investors simply buy low and sell high. They buy A, which offers a high return relative to the CAPM, and they sell B, whose return is too low relative to the CAPM. In a perfect world, these unconstrained investors would bid up the price of A until it no longer has any excess returns. And they would sell B until its returns reach a fair level relative to the CAPM. In this perfect world, the risk anomaly disappears.

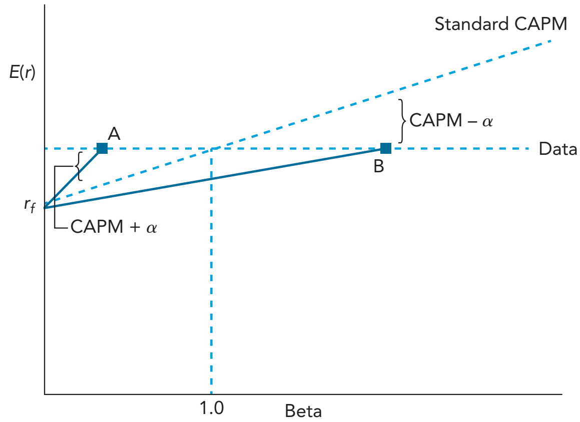


Figure 3.14

Now consider a long-only investor subject to tracking error constraints that place limits on how much she can deviate from benchmark (see Equation (3.3)). This investor cannot short. This investor does *not* invest in A, even though A is offering high returns relative to the CAPM. The returns of A are higher than the CAPM predicts. Stock A would even perform in line with the market. But, by investing in A, she takes on significant tracking error.

What about stock B? Stock B has negative alpha. To make money, she needs to short stock B, and she cannot do so. The best that she can do is to avoid buying stock B, thus making at most a small active bet relative to the market portfolio. The tracking error constraint also limits the underweight position in stock B that she can hold. If the "Data" line is in fact just slightly upward sloping rather than perfectly flat, then she actually has an incentive to buy B rather than sell it short because B could outperform the market.

Thus, the use of tracking error with these benchmarks makes it hard to bet against low volatility or low beta. Tracking error is a binding constraint for GM Asset Management. It is also a binding constraint for most institutional asset owners. One obvious solution is to change the benchmarks, and there are certainly more appropriate factor benchmarks available. But changing benchmarks at GM is a lengthy process requiring approval of the investment committee. It opens up a broader issue of how all benchmarks "depend on funded status and on the health of the parent," as Scott explains.

Frazzini, Kabiller, and Pedersen (2012) even argue that low-risk factors play a part in explaining the superior performance of Berkshire Hathaway—a company well known for its ability to go against the crowd and avoid common agency issues. They

[30] Black (1972) was the first to develop a theory of the CAPM for when investors cannot lever. Frazzini and Pedersen (2011) apply a leverage-constraint story to explain the low-beta anomaly.

[31] For an academic reference, see Han and Kumar (2011). Taking 13-F filings as of June 30, 2012 on Russell 1000 holdings, Martingale calculates that institutions hold 46.5% low-risk stocks and 53.5% high-risk stocks compared to a balanced 50%/50% split.

[32] See Greenwood et. al. (2010) and Baker, Bradley, and Wurgler (2011).

find that Buffett's alpha declines from 12.5% from 1976 to 2011 using the Fama-French and momentum benchmark we've been using in this chapter to 11.1% when including the *BAB* factor. If they add another factor measuring the underlying quality of companies, Buffett's alpha falls to 7.0%. So some of Buffett's investing prowess is due to Buffett selecting stocks with low risk, but most of Buffett's investment prowess comes from ferreting out gems with high underlying quality—true skill that is unrelated to just holding low-volatility stocks.

Preferences

If asset owners simply have a preference for high-volatility and high-beta stocks, then they bid up these stocks until they have low returns. Conversely, these investors shun safe stocks—stocks with low volatility and low betas—leading to low prices and high returns for these shares. Thus, "hopes and dreams" preferences, where the hopes and dreams are represented by high-volatility and high-beta stocks, could explain the risk anomaly.[33]

Hou and Loh (2012) comprehensively examine many explanations of the low volatility anomaly. They arrange their explanations into three broad groups: (i) lottery preference, (ii) market frictions including illiquidity, and (iii) "other," which is a broad category that includes uncertainty, short-sales constraints, financial distress, investor inattention, growth options, earnings shocks, and other variables. Hou and Loh find that when individual explanations are taken alone, each explains less than one-tenth of the volatility anomaly. But taken as groups, the most promising explanation is lottery preferences. When individual lottery preference stories are taken together, they explain close to half of the low volatility puzzle. But close to half of the puzzle remains unexplained.

Agents disagreeing with each other (heterogeneous preferences) combined with the inability to short could also account for some of the risk anomaly. Hong and Sraer (2012) show that when disagreement is low and everyone takes long-only positions, the CAPM holds. But when disagreement is high, some agents want to sell short and they cannot. High beta stocks become overpriced. Large enough disagreement causes the relation between beta and returns to be downward sloping.[34]

3.6 GM ASSET MANAGEMENT AND MARTINGALE REDUX

Martingale's low volatility strategy is attractive compared to the market portfolio. It delivers alpha relative to the Russell 1000 benchmark of 1.50% per year. Adjusting the Russell 1000 for risk increases that alpha to 3.44% per year. Alpha is all about the benchmark. What if we changed the benchmark of Martingale's strategy to be the low volatility strategy itself? Then, there wouldn't be any alpha of course, as alpha morphs into the benchmark (or beta, as some in industry like to call it). This is not just philosophical—GM Asset Management might be in a position to internally do low volatility strategies. But low-risk strategies appear to have significant alpha relative to standard market capitalization benchmarks and sophisticated factor benchmarks that control for risk using dynamic value-growth and momentum factors along with the market portfolio.

Yet alpha is not the only consideration for GM Asset Management. Martingale's alpha comes with high tracking error relative to the Russell 1000 benchmark. In fact, the ubiquitous tracking error constraints employed in the asset management industry may partly give rise to the risk anomaly in the first place.

Will the risk anomaly persist? I am hoping that it goes away as soon as possible, and I have a large academic stake in this debate. As much as I enjoy seeing new explanations being proposed (including some of my own), the risk anomaly is an enigma. If it does disappear, then the low-risk trades already put on by the smart money will payoff handsomely—low-volatility or low-beta stocks have returns that are too high and prices that are too low. Capital should be drawn to these stocks, driving up their prices and removing the anomalous returns. If that happened, current low-risk anomaly investors would enjoy large capital gains.

But I doubt this will happen. Low-volatility strategies are far from predominant, as most institutional investors appear to be underweight low-risk stocks, More fundamentally, the fact that we see the risk anomaly in many markets—U.S. and international, stocks, bonds, commodities, foreign exchange, and derivatives—suggests that the effect is pervasive and requires a deep explanation. As Greenwood says, the low-risk anomaly is the mother of all inefficiencies.

[33] For stories along these lines, see Boyer, Mitton, and Vorkink (2010), Bali, Cakici, and Whitelaw (2011), and Ilmanen (2012).

[34] See also Jiang, Xu, and Yao (2009) for the relation between earnings uncertainty and low volatility.

4

Portfolio Construction

Learning Objectives

After completing this reading you should be able to:

- Distinguish among the inputs to the portfolio construction process.
- Evaluate the motivation for and the methods used for refining alphas in the implementation process.
- Describe neutralization and the different approaches used for refining alphas to be neutral.
- Describe the implications of transaction costs on portfolio construction.
- Describe practical issues in portfolio construction, including the determination of an appropriate risk aversion, aversions to specific risks, and proper alpha coverage.
- Describe portfolio revisions and rebalancing, and analyze the tradeoffs between alpha, risk, transaction costs, and time horizon.
- Determine the optimal no-trade region for rebalancing with transaction costs.
- Evaluate the strengths and weaknesses of the following portfolio construction techniques: screens, stratification, linear programming, and quadratic programming.
- Describe dispersion, explain its causes, and describe methods for controlling forms of dispersion.

Excerpt is Chapter 14 of Active Portfolio Management: A Quantitative Approach for Providing Superior Returns and Controlling Risk, *Second Edition, by Richard C. Grinold and Ronald N. Kahn.*

4.1 INTRODUCTION

Implementation is the efficient translation of research into portfolios. Implementation is not glamorous, but it is important. Good implementation can't help poor research, but poor implementation can foil good research. A manager with excellent information and faulty implementation can snatch defeat from the jaws of victory.

Implementation includes both portfolio construction, the subject of this chapter, and trading. This chapter will take a manager 's investment constraints (e.g., no short sales) as given and build the best possible portfolio subject to those limitations. It will assume the standard objective: maximizing active returns minus an active risk penalty. This chapter will also take transactions costs as just an input to the portfolio construction problem.

Portfolio construction requires several inputs: the current portfolio, alphas, covariance estimates, transactions cost estimates, and an active risk aversion. Of these inputs, we can measure only the current portfolio with near certainty. The alphas, covariances, and transactions cost estimates are all subject to error. The alphas are often unreasonable and subject to hidden biases. The covariances and transactions costs are noisy estimates; we hope that they are unbiased, but we know that they are not measured with certainty. Even risk aversion is not certain. Most active managers will have a target level of active risk that we must make consistent with an active risk aversion.

Implementation schemes must address two questions. First, what portfolio would we choose given inputs (alpha, covariance, active risk aversion, and transactions costs) known without error? Second, what procedures can we use to make the portfolio construction process robust in the presence of unreasonable and noisy inputs? How do you handle perfect data, and how do you handle less than perfect data?

How to handle perfect data is the easier dilemma. With no transactions costs, the goal is to maximize value added within any limitations on the manager's behavior imposed by the client. Transactions costs make the problem more difficult. We must be careful to compare transactions costs incurred at a point in time with returns and risk realized over a period of time.

This chapter will mainly focus on the second question, how to handle less than perfect data. Many of the procedures used in portfolio construction are, in fact, indirect methods of coping with noisy data. With that point of view, we hope to make portfolio construction more efficient by directly attacking the problem of imperfect or "noisy" inputs.

Several points emerge in this chapter:

- Implementation schemes are, in part, safeguards against poor research.
- With alpha analysis, the alphas can be adjusted so that they are in line with the manager 's desires for risk control and anticipated sources of value added.
- Portfolio construction techniques include screening, stratified sampling, linear programming, and quadratic programming. Given sufficiently accurate risk estimates, the quadratic programming technique most consistently achieves high value added.
- For most active institutional portfolio managers, building portfolios using alternative risk measures greatly increases the effort (and the chance of error) without greatly affecting the result.
- Managers running separate accounts for multiple clients can control dispersion, but cannot eliminate it.

Let's start with the relationship between the most important input, alpha, and the output, the revised portfolio.

4.2 ALPHAS AND PORTFOLIO CONSTRUCTION

Active management should be easy with the right alphas. Sometimes it isn't. Most active managers construct portfolios subject to certain constraints, agreed upon with the client. For example, most institutional portfolio managers do not take short positions and limit the amount of cash in the portfolio. Others may restrict asset coverage because of requirements concerning liquidity, self-dealing, and so on. These limits can make the portfolio less efficient, but they are hard to avoid.

Managers often add their own restrictions to the process. A manager may require that the portfolio be neutral across economic sectors or industries. The manager may limit individual stock positions to ensure diversification of the active bets. The manager may want to avoid any position based on a forecast of the benchmark portfolio's performance. Managers often use such restrictions to make portfolio construction more robust.

There is another way to reach the same final portfolio: simply adjust the inputs. We can always replace a very sophisticated (i.e., complicated) portfolio construction procedure that leads to active holdings $\mathbf{h}^*_{PA}$, active risk ψ^*_P, and an ex ante information ratio IR with a direct unconstrained mean/variance optimization

Table 4.1

Stock	Index Weight	Alpha	Optimal Holding	Constrained Optimal Holding	Modified Alpha
American Express	2.28%	−3.44%	−0.54%	0.00%	−1.14%
AT&T	4.68%	1.38%	6.39%	6.18%	0.30%
Chevron	6.37%	0.56%	7.41%	7.05%	0.11%
Coca-Cola	3.84%	−2.93%	−2.22%	0.00%	−0.78%
Disney	3.94%	1.77%	5.79%	5.85%	0.60%
Dow Chemical	5.25%	0.36%	5.78%	6.07%	0.22%
DuPont	4.32%	−1.50%	1.54%	1.67%	−0.65%
Eastman Kodak	3.72%	0.81%	4.07%	4.22%	0.14%
Exxon	5.60%	−0.10%	4.57%	4.39%	−0.19%
General Electric	7.84%	−2.80%	0.53%	0.92%	−1.10%
General Motors	2.96%	−2.50%	1.93%	1.96%	−0.52%
IBM	4.62%	−2.44%	3.24%	3.54%	−0.51%
International Paper	6.11%	−0.37%	5.73%	6.15%	0.01%
Johnson & Johnson	4.63%	2.34%	7.67%	7.71%	0.66%
McDonalds	4.47%	0.86%	5.07%	4.98%	0.14%
Merck	3.98%	0.80%	4.72%	4.78%	0.20%
3M	9.23%	3.98%	17.95%	14.23%	0.91%
Philip Morris	7.07%	0.71%	7.82%	7.81%	0.12%
Procter & Gamble	4.92%	1.83%	6.99%	6.96%	0.44%
Sears	4.17%	0.69%	5.57%	5.54%	0.35%

using a modified set of alphas and the appropriate level of risk aversion.[1] The modified alphas are

$$\boldsymbol{\alpha}' = \left(\frac{\text{IR}}{\psi_P^*}\right) \cdot \mathbf{V} \cdot \mathbf{h}_{PA}^* \tag{4.1}$$

and the appropriate active risk aversion is

$$\lambda_A' = \frac{\text{IR}}{2 \cdot \psi_P^*} \tag{4.2}$$

Table 4.1 illustrates this for Major Market Index stocks as of December 1992. We assign each stock an alpha (chosen randomly in this example), and first run an unconstrained optimization of risk-adjusted active return (relative to the Major Market Index) using an active risk aversion of 0.0833. Table 4.1 shows the result. The unconstrained optimization sells American Express and Coca-Cola short, and invests almost 18 percent of the portfolio in 3M. We then add constraints; we disallow short sales and require that portfolio holdings cannot exceed benchmark holdings by more than 5 percent. This result is also displayed in Table 4.1. The optimal portfolio no longer holds American Express or Coca-Cola at all, and the holding of 3M moves to exactly 5 percent above the benchmark holding. The other positions also adjust.

This constrained optimization corresponds to an unconstrained optimization using the same active risk aversion of 0.0833 and the modified alphas displayed in the last column of Table 4.1. We derive these using Equations (4.1) and (4.2). These modified alphas are pulled in toward zero relative to the original alphas, as we would expect, since the constraints moved the optimal

[1] The simple procedure maximizes $\mathbf{h}T_{PA} \cdot \boldsymbol{\alpha}' - \lambda_A' \cdot \mathbf{h}T_{PA} \cdot \mathbf{V} \cdot \mathbf{h}_{PA}$. The first-order conditions for this problem are $\boldsymbol{\alpha}' = 2 \cdot \lambda_A' \cdot \mathbf{V} \cdot \mathbf{h}_{PA}^*$. Equations (4.1) and (4.2) ensure that $\mathbf{h}_{PA}$ will satisfy the first-order conditions. Note that we are explicitly focusing portfolio construction on active return and risk, instead of residual return and risk. Without benchmark timing, these perspectives are identical.

portfolio closer to the benchmark. The original alphas have a standard deviation of 2.00 percent, while the modified alphas have a standard deviation of 0.57 percent.

We can replace any portfolio construction process, regardless of its sophistication, by a process that first refines the alphas and then uses a simple unconstrained mean/variance optimization to determine the active positions.

This is not an argument against complicated implementation schemes. It simply focuses our attention on a reason for the complexity. If the implementation scheme is, in part, a safeguard against unrealistic or unreasonable inputs, perhaps we can, more fruitfully, address this problem directly. A direct attack calls for either refining the alphas (preprocessing) or designing implementation procedures that explicitly recognize the procedure's role as an "input moderator." The next section discusses preprocessing of alphas.

4.3 ALPHA ANALYSIS

We can greatly simplify the implementation procedure if we ensure that our alphas are consistent with our beliefs and goals. Here we will outline some procedures for refining alphas that can simplify the implementation procedure, and explicitly link our refinement in the alphas to the desired properties of the resulting portfolios. We begin with the standard data screening procedures of scaling and trimming.[2]

Scale the Alphas

Alphas have a natural structure: $\alpha = \text{volatility} \cdot \text{IC} \cdot \text{score}$. This structure includes a natural scale for the alphas. We expect the information coefficient (IC) and residual risk (volatility) for a set of alphas to be approximately constant, with the score having mean 0 and standard deviation 1 across the set. Hence the alphas should have mean 0 and standard deviation, or *scale*, of $\text{Std}\{\alpha\} \sim \text{volatility} \cdot \text{IC}$.[3] An information coefficient of 0.05 and a typical residual risk of 30 percent would lead to an alpha scale of 1.5 percent. In this case, the mean alpha would be 0, with roughly two-thirds of the stocks having alphas between −1.5 percent and +1.5 percent and roughly 5 percent of the stocks having alphas larger than +3.0 percent or less than −3.0 percent. In Table 4.1, the original alphas have a standard deviation of 2.00 percent and the modified alphas have a standard deviation of 0.57 percent. This implies that the constraints in that example effectively shrank the IC by 62 percent, a significant reduction. There is value in noting this explicitly, rather than hiding it under a rug of optimizer constraints.

The scale of the alphas will depend on the information coefficient of the manager. If the alphas input to portfolio construction do not have the proper scale, then rescale them.

Trim Alpha Outliers

The second refinement of the alphas is to trim extreme values. Very large positive or negative alphas can have undue influence. Closely examine all stocks with alphas greater in magnitude than, say, three times the scale of the alphas. A detailed analysis may show that some of these alphas depend upon questionable data and should be ignored (set to zero), while others may appear genuine. Pull in these remaining genuine alphas to three times scale in magnitude.

A second and more extreme approach to trimming alphas is to force[4] them into a normal distribution with benchmark alpha equal to 0 and the required scale factor. Such an approach is extreme because it typically utilizes only the ranking information in the alphas and ignores the size of the alphas. After such a transformation, you must recheck benchmark neutrality and scaling.

Neutralization

Beyond scaling and trimming, we can remove biases or undesirable bets from our alphas. We call this process *neutralization*. It has implications, not surprisingly, in terms of both alphas and portfolios.

Benchmark neutralization means that the benchmark has 0 alpha. If our initial alphas imply an alpha for the benchmark,

[2] Because of their simplicity, we treat scaling and trimming first. However, when we implement alpha analysis, we impose scaling and trimming as the final step in the process.

[3] There is a related approach to determining the correct scale that uses the information ratio instead of the information coefficient. This approach calculates the information ratio implied by the alphas and scales them, if necessary, to match the manager's ex ante information ratio. The information ratio implied by the alphas is $\text{IR}_0 = \sqrt{\boldsymbol{\alpha}^T \cdot \mathbf{V}^{-1} \cdot \boldsymbol{\alpha}}$. We can calculate this quickly by running an optimization with unrestricted cash holdings, no constraints, no limitations on asset holdings, and an active risk aversion of 0.5. The optimal active portfolio is $\mathbf{h}_{PA}^* = \mathbf{V}^{-1} \cdot \boldsymbol{\alpha}$, and the optimal portfolio alpha is $(\text{IR}_0)^2$. If IR is the desired ex ante information ratio, we can rescale the alphas by a factor (IR/IR_0).

[4] Suppose that $h_{B,n}$ is the benchmark weight for asset n. Assume for convenience that the assets are ordered so that $\alpha_1 \leq \alpha_2 \leq \alpha_3$, etc. Then define $p_1 = 0.5 \cdot h_{B,1}$ and for $n \geq 2$, $p_n = p_{n-1} + 0.5 \cdot (h_{B,n-1} + h_{B,n})$. We have $0 < p_1 < p_2 < \cdots < p_{N-1} < p_N < 1$. Find the normal variate z_n that satisfies $p_n = \Phi\{z_n\}$, where Φ is the cumulative normal distribution. We can use the z variables as alphas, after adjustments for location and scale.

the neutralization process recenters the alphas to remove the benchmark alpha. From the portfolio perspective, benchmark neutralization means that the optimal portfolio will have a beta of 1, i.e., the portfolio will not make any bet on the benchmark.

Neutralization is a sophisticated procedure, but it isn't uniquely defined. We can achieve even benchmark neutrality in more than one way. This is easy to see from the portfolio perspective: We can choose many different portfolios to hedge out any active beta.

As a general principle, we should consider a priori how to neutralize our alphas. The choices will include benchmark, cash, industry, and factor neutralization. Do our alphas contain any information distinguishing one industry from another? If not, then industry-neutralize. The a priori approach works better than simply trying all possibilities and choosing the best performer.

Benchmark- and Cash-Neutral Alphas

The first and simplest neutralization is to make the alphas benchmark-neutral. By definition, the benchmark portfolio has 0 alpha, although the benchmark may experience exceptional return. Setting the benchmark alpha to 0 ensures that the alphas are benchmark-neutral and avoids benchmark timing.

In the same spirit, we may also want to make the alphas cash-neutral; i.e., the alphas will not lead to any active cash position. It is possible to make the alphas both cash- and benchmark-neutral.

Table 4.2 displays the modified alphas from Table 4.1 and shows how they change when we make them benchmark-neutral. In this example, the benchmark alpha is only 1.6 basis points, so subtracting $\beta_n \cdot \alpha_B$ from each modified alpha does not change the alpha very much. We have shifted the alpha of the benchmark Major Market Index from 1.6 basis points to 0. This small change in alpha is consistent with the observation that the optimal portfolio before benchmark neutralizing had a beta very close to 1.

Risk-Factor-Neutral Alphas

The multiple-factor approach to portfolio analysis separates return along several dimensions. A manager can identify each of those dimensions as either a source of risk or a source of value added. By this definition, the manager does not have any ability to forecast the risk factors. Therefore, he or she should neutralize the alphas against the risk factors. The neutralized alphas will include only information on the factors the manager can forecast, along with specific asset information. Once neutralized, the alphas of the risk factors will be 0.

Table 4.2

Stock	Beta	Modified Alpha	Modified Benchmark-Neutral Alpha
American Express	1.21	−1.14%	−1.16%
AT&T	0.96	0.30%	0.29%
Chevron	0.46	0.11%	0.10%
Coca Cola	0.96	−0.78%	−0.79%
Disney	1.23	0.60%	0.58%
Dow Chemical	1.13	0.22%	0.20%
DuPont	1.09	−0.65%	−0.67%
Eastman Kodak	0.60	0.14%	0.13%
Exxon	0.46	−0.19%	−0.20%
General Electric	1.30	−1.10%	−1.12%
General Motors	0.90	−0.52%	−0.53%
IBM	0.64	−0.51%	−0.52%
International Paper	1.18	0.01%	−0.01%
Johnson & Johnson	1.13	0.66%	0.64%
McDonalds	1.06	0.14%	0.12%
Merck	1.06	0.20%	0.18%
3M	0.74	0.91%	0.90%
Philip Morris	0.94	0.12%	0.10%
Procter & Gamble	1.00	0.44%	0.42%
Sears	1.05	0.35%	0.33%

For example, a manager can ensure that her portfolios contain no active bets on industries or on a size factor. Here is one simple approach to making alphas industry-neutral: Calculate the (capitalization-weighted) alpha for each industry, then subtract the industry average alpha from each alpha in that industry.

We can modify the alphas to achieve desired active common-factor positions and to isolate the part of the alpha that does not influence the common-factor positions.

4.4 TRANSACTIONS COSTS

Up to this point, the struggle has been between alpha and active risk. Any klutz can juggle two rubber chickens. The juggling becomes complicated when the third chicken enters the performance. In portfolio construction, that third rubber chicken is *transactions costs*, the cost of moving from one portfolio to

another. It has been said that accurate estimation of transactions costs is just as important as accurate forecasts of exceptional return. That is an overstatement,[5] but it does point out the crucial role transactions costs play.

In addition to complicating the portfolio construction problem, transactions costs have their own inherent difficulties. We will see that transactions costs force greater precision on our estimates of alpha. We will also confront the complication of comparing transactions costs at a point in time with returns and risk which occur over an investment horizon.

When we consider only alphas and active risk in the portfolio construction process, we can offset any problem in setting the scale of the alphas by increasing or decreasing the active risk aversion. Finding the correct trade-off between alpha and active risk is a one-dimensional problem. By turning a single knob, we can find the right balance. Transactions costs make this a two-dimensional problem. The trade-off between alpha and active risk remains, but now there is a new trade-off between the alpha and the transactions costs. We therefore must be precise in our choice of scale, to correctly trade off between the hypothetical alphas and the inevitable transactions costs.

The objective in portfolio construction is to maximize risk-adjusted annual active return. Rebalancing incurs transactions costs at that point in time. To contrast transactions costs incurred at that time with alphas and active risk expected over the next year requires a rule to allocate the transactions costs over the one-year period. We must amortize the transactions costs to compare them to the annual rate of gain from the alpha and the annual rate of loss from the active risk. The rate of amortization will depend on the anticipated holding period.

An example will illustrate this point. We will assume perfect certainty and a risk-free rate of zero; and we will start and end invested in cash. Stock 1's current price is $100. The price of stock 1 will increase to $102 in the next 6 months and then remain at $102. Stock 2's current price is also $100. The price of stock 2 will increase to $108 over the next 24 months and then remain at $108. The cost of buying and selling each stock is $0.75. The annual alpha for both stock 1 and stock 2 is 4 percent. To contrast the two situations more clearly, let's assume that in 6 months, and again in 12 months and in 18 months, we can find another stock like stock 1.

The sequence of 6-month purchases of stock 1 and its successors will each net a $2.00 profit before transactions costs. There will be transactions costs (recall that we start and end with cash) of $0.75, $1.50, $1.50, $1.50, and $0.75 at 0, 6, 12, 18, and 24 months, respectively. The total trading cost is $6, the gain on the shares is $8, the profit over 2 years is $2, and the annual percentage return is 1 percent.

With stock 2, over the 2-year period we will incur costs of $0.75 at 0 and 24 months. The total cost is $1.50, the gain is $8, the profit is $6.50, and the annual percentage return is 3.25 percent.

With the series of stock 1 trades, we realize an annual alpha of 4 percent and an annualized transactions cost of 3 percent. With the single deal in stock 2, we realize an annual alpha of 4 percent and an annualized transactions cost of 0.75 percent. For a 6-month holding period, we double the round-trip transactions cost to get the annual transactions cost, and for a 24-month holding period, we halve the round-trip transactions cost to get the annual transactions cost. There's a general rule here:

> The annualized transactions cost is the round-trip cost divided by the holding period in years.

For the remainder of this chapter, we will assume that we know the cost for each anticipated trade.

4.5 PRACTICAL DETAILS

Before proceeding further in our analysis of portfolio construction, we should review some practical details concerning this process. First, how do we choose a risk aversion parameter?

We can find an optimality relationship between the information ratio, the risk aversion, and the optimal active risk. That result is displayed here, translated from residual to active return and risk,

$$\lambda_A = \frac{\text{IR}}{2 \cdot \psi_P} \tag{4.3}$$

The point is that we have more intuition about our information ratio and our desired amount of active risk. Hence, we can use Equation (4.3) to back out an appropriate risk aversion. If our information ratio is 0.5, and we desire 5 percent active risk, we should choose an active risk aversion of 0.05. Note that we must be careful to verify that our optimizer is using percents and not decimals.

A second practical matter concerns aversion to specific as opposed to common-factor risk. Several commercial optimizers utilize this decomposition of risk to allow differing aversions to these different sources of risk:

$$U = \alpha_P - (\lambda_{A,CF} \cdot \Psi^2_{P,CF} + \lambda_{A,SP} \cdot \Psi^2_{P,SP}) \tag{4.4}$$

An obvious reaction here is, "Risk is risk, why would I want to avoid one source of risk more than another?" This is a useful

[5] Perfect information regarding returns is much more valuable than perfect information regarding transactions costs. The returns are much less certain than the transactions costs. Accurate estimation of returns reduces uncertainty much more than accurate estimation of transactions costs.

sentiment to keep in mind, but there are at least two reasons to consider implementing a higher aversion to specific risk. First, since specific risk arises from bets on specific assets, a high aversion to specific risk reduces bets on any one stock. In particular, this will reduce the size of your bets on the (to be determined) biggest losers. Second, for managers of multiple portfolios, aversion to specific risk can help reduce dispersion. This will push all those portfolios toward holding the same names.

The final practical details we will cover here concern alpha coverage. First, what happens if we forecast returns on stocks that are not in the benchmark? We can always handle that by expanding the benchmark to include those stocks, albeit with zero weight. This keeps stock n in the benchmark, but with no weight in determining the benchmark return or risk. Any position in stock n will be an active position, with active risk correctly handled.

What about the related problem, a lack of forecast returns for stocks in the benchmark? For stock-specific alphas, we can use the following approach.

Let N_1 represent the collection of stocks with forecasts, and N_0 the stocks without forecasts. The value-weighted fraction of stocks with forecasts is

$$H\{N_1\} = \sum_{n \in N_1} h_{B,n} \tag{4.5}$$

The average alpha for group N_1 is

$$\alpha\{N_1\} = \frac{\sum_{n \in N_1} h_{B,n} \cdot \alpha_n}{H\{N_1\}} \tag{4.6}$$

To round out the set of forecasts, set $\alpha_n^* = \alpha_n - \alpha\{N_1\}$ for stocks in N_1 and $\alpha_n^* = 0$ for stocks in N_0. These alphas are benchmark-neutral. Moreover, the stocks we did not cover will have a zero, and therefore neutral, forecast.

4.6 PORTFOLIO REVISIONS

How often should you revise your portfolio? Whenever you receive new information. That's the short answer. If a manager knows how to make the correct trade-off between expected active return, active risk, and transactions costs, frequent revision will not present a problem. If the manager has human failings, and is not sure of his or her ability to correctly specify the alphas, the active risk, and the transactions costs, then the manager may resort to less frequent revision as a safeguard.

Consider the unfortunate manager who underestimates transactions costs, makes large changes in alpha estimates very frequently, and revises his portfolio daily. This manager will churn the portfolio and suffer higher than expected transactions costs and lower than expected alpha. A crude but effective cure is to revise the portfolio less frequently.

More generally, even with accurate transactions costs estimates, as the horizon of the forecast alphas decreases, we expect them to contain larger amounts of noise. The returns themselves become noisier with shorter horizons. Rebalancing for very short horizons would involve frequent reactions to noise, not signal. But the transactions costs stay the same, whether we are reacting to signal or noise.

This trade-off between alpha, risk, and costs is difficult to analyze because of the inherent importance of the horizon. We expect to realize the alpha over some horizon. We must therefore amortize the transactions costs over that horizon.

We can capture the impact of new information, and decide whether to trade, by comparing the marginal contribution to value added for stock n, $MCVA_n$, to the transactions costs. The marginal contribution to value added shows how value added, as measured by risk-adjusted alpha, changes as the holding of the stock is increased, with an offsetting decrease in the cash position. As our holding in stock n increases, α_n measures the effect on portfolio alpha. The change in value added also depends upon the impact (at the margin) on active risk of adding more of stock n. The stock's marginal contribution to active risk, $MCAR_n$, measures the rate at which active risk changes as we add more of stock n. The loss in value added due to changes in the level of active risk will be proportional to $MCAR_n$. Stock n's marginal contribution to value added depends on its alpha and marginal contribution to active risk, in particular:

$$MCVA_n = \alpha_n - 2 \cdot \lambda_A \cdot \psi \cdot MCAR_n \tag{4.7}$$

Let PC_n be the purchase cost and SC_n the sales cost for stock n. For purposes of illustration, we take $PC_n = 0.50$ percent and $SC_n = 0.75$ percent. If the current portfolio is optimal,[6] then the marginal contribution to value added for stock n should be less than the purchase cost. If it exceeded the purchase cost, say at 0.80 percent, then a purchase of stock n would yield a net benefit of 0.80 percent $-$ 0.50 percent $=$ 0.30 percent. Similarly the marginal contribution to value added must be greater than the negative of the sales cost. If it were -1.30 percent, then we could decrease our holding of stock n and save 1.30 percent at the margin. The cost would be the 0.75 percent transactions cost, for a net benefit of 1.30 percent $-$ 0.75 percent $=$ 0.55 percent.

[6] Assuming no limitations on holdings, no limitations on the cash position, and no additional constraints. Aficionados will realize that this analysis becomes more complicated, but not essentially different, if we include these additional constraints.

This observation allows us to put a band around the alpha for each stock. As long as the alpha stays within that band, the portfolio will remain optimal, and we should not react to new information. The bandwidth is the total of the sale plus purchase costs, 0.50 percent + 0.75 percent = 1.25 percent in our example. If we just purchased a stock, its marginal contribution to value added will equal its purchase cost. We are at the upper end of the band. Any increase in alpha would lead to further purchases. The alpha would have to decrease by 1.25 percent before we would consider selling the stock. The situation before new information arrives is

$$-SC_n \leq MCVA_n \leq PC_n \tag{4.8}$$

or, using Equation (4.7),

$$2 \cdot \lambda_A \cdot \psi \cdot MCAR_n - SC_n \leq \alpha_n \leq PC_n + 2 \cdot \lambda_A \cdot \psi \cdot MCAR_n \tag{4.9}$$

This analysis has simplified the problem by subsuming the amortization horizon into the costs SC and PC. To fully treat the issue of when to rebalance requires analyzing the dynamic problem involving alphas, risks, and costs over time. There are some useful results from this general treatment, in the very simple case of one or two assets.

Leland (1996) solves the asset allocation problem of rebalancing around an optimal stock/bond allocation. Let's assume that the optimal allocation is 60/40. Assuming linear transactions costs and a utility function penalizing active variance (relative to the optimal allocation) and transactions costs *over time,* Leland shows that the optimal strategy involves a no-trade region around the 60/40 allocation. If the portfolio moves outside that region, the optimal strategy is to trade back to the boundary. Trading only to the boundary, not to the target allocation, cuts the turnover and transactions costs roughly in half, with effectively no change in risk over time. The size of the no-trade region depends on the transactions costs, the risk aversion, and the expected return and risk of stocks and bonds. Obviously, changing the size of the no-trade region will change the turnover for the strategy.

This result concerns a problem that is much simpler than our general active portfolio management problem: The solved problem is one-dimensional and does not involve the flow of information (the target allocation is static). Still, it is useful in motivating rebalancing rules driven not purely by the passage of time (e.g., monthly or quarterly rebalancing), but rather by the portfolio's falling outside certain boundaries.

Another approach to the dynamic problem utilizes information horizon analysis. Here we apply trading rules like Equation (4.9)

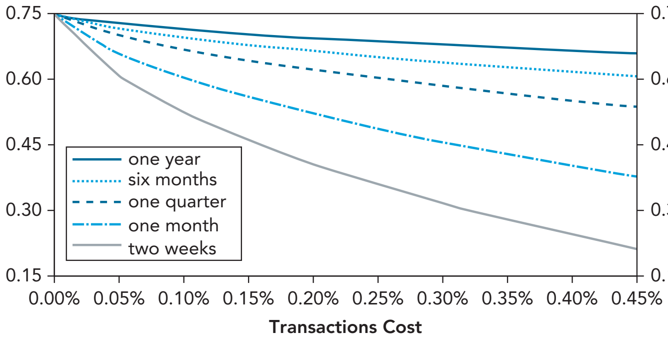


Figure 4.1 **After-cost information ratio for various half-lives.**

in the dynamic case of trading one position only, over an indefinite future,[7] with information characterized by an information horizon. Figure 4.1 shows how the after-cost information ratio declines as a function of both the (one-way) cost and the half-life of the signals. Two effects are at work. First, when we trade, we pay the costs. Second, and more subtle, the transactions costs makes us less eager; we lose by intimidation.

4.7 TECHNIQUES FOR PORTFOLIO CONSTRUCTION

There are as many techniques for portfolio construction as there are managers. Each manager adds a special twist. Despite this personalized nature of portfolio construction techniques, there are four generic classes of procedures that cover the vast majority of institutional portfolio management applications:[8]

- Screens
- Stratification
- Linear programming
- Quadratic programming

Before we examine these procedures in depth, we should recall our criteria. We are interested in high alpha, low active risk, and

[7] There is a pleasant symmetry in this approach. Conventional portfolio optimization considers lots of assets in a one-period framework; we are considering one-asset (position) in a multiple-period framework.

[8] The techniques we review successfully handle monthly or quarterly rebalancing of portfolios of up to 1000 assets and asset universes that can exceed 10,000 for international investing. Later, we will discuss nonlinear programming and stochastic optimization, whose applications are generally limited to asset allocation schemes involving few (less than 25) asset classes and long planning horizons.

low transactions costs. Our figure of merit is value added less transactions costs:

$$\alpha_P - \lambda_A \cdot \psi_P^2 - TC \tag{4.10}$$

We will see how each of these procedures deals with these three aspects of portfolio construction.

Screens

Screens are simple. Here is a screen recipe for building a portfolio from scratch:

1. Rank the stocks by alpha.
2. Choose the first 50 stocks (for example).
3. Equal-weight (or capitalization-weight) the stocks.

We can also use screens for rebalancing. Suppose we have alphas on 200 stocks (the followed list). Divide the stocks into three categories: the top 40, the next 60, and the remaining 100. Put any stock in the top 40 on the buy list, any stock in the bottom 100 on the sell list, and any stock in the middle 60 on the hold list. Starting with the current 50-stock portfolio, buy any stocks that are on the buy list but not in the portfolio. Then sell any assets that are in the portfolio and on the sell list. We can adjust the numbers 40, 60, and 100 to regulate turnover.

Screens have several attractive features. There is beauty in simplicity. The screen is easy to understand, with a clear link between cause (membership on a buy, sell, or hold list) and effect (membership in the portfolio). The screen is easy to computerize; it might be that mythical computer project that can be completed in two days! The screen is robust. Notice that it depends solely on ranking. Wild estimates of positive or negative alphas will not alter the result.

The screen enhances alphas by concentrating the portfolio in the high-alpha stocks. It strives for risk control by including a sufficient number of stocks (50 in the example) and by weighting them to avoid concentration in any single stock. Transactions costs are limited by controlling turnover through judicious choice of the size of the buy, sell, and hold lists.

Screens also have several shortcomings. They ignore all information in the alphas apart from the rankings. They do not protect against biases in the alphas. If all the utility stocks happen to be low in the alpha rankings, the portfolio will not include any utility stocks. Risk control is fragmentary at best. In our consulting experience, we have come across portfolios produced by screens that were considerably more risky than their managers had imagined. In spite of these significant shortcomings, screens are a very popular portfolio construction technique.

Stratification

Stratification is glorified screening. The term *stratification* comes from statistics. In statistics, stratification guards against sample bias by making sure that the sample population is representative of the total population as it is broken down into distinct subpopulations. The term is used very loosely in portfolio construction. When a portfolio manager says he uses stratified sampling, he wants the listener to (1) be impressed and (2) ask no further questions.

The key to stratification is splitting the list of followed stocks into categories. These categories are generally exclusive. The idea is to obtain risk control by making sure that the portfolio has a representative holding in each category. As a typical example, let's suppose that we classify stocks into 10 economic sectors and also classify the stocks in each sector by size: big, medium, and small. Thus, we classify all stocks into 30 categories based on economic sector and size. We also know the benchmark weight in each of the 30 categories.

To construct a portfolio, we mimic the screening exercise within each category. We rank the stocks by alpha and place them into buy, hold, and sell groups within each category in a way that will keep the turnover reasonable. We then weight the stocks so that the portfolio's weight in each category matches the benchmark's weight in that category. Stratification ensures that the portfolio matches the benchmark along these important dimensions.

The stratification scheme has the same benefits as screening, plus some. It is robust. Improving upon screening, it ignores any biases in the alphas across categories. It is somewhat transparent and easy to code. It has the same mechanism as screening for controlling turnover.

Stratification retains some of the shortcomings of a screen. It ignores some information, and does not consider slightly over-weighting one category and underweighting another. Often, little substantive research underlies the selection of the categories, and so risk control is rudimentary. Chosen well, the categories can lead to reasonable risk control. If some important risk dimensions are excluded, risk control will fail.

Linear Programming

A linear program (LP) is space-age stratification. The linear programming approach[9] characterizes stocks along dimensions of risk, e.g., industry, size, volatility, and beta. The linear program

[9] A linear program is a useful tool for a variety of portfolio management applications. The application described here is but one of those applications.

does not require that these dimensions distinctly and exclusively partition the stocks. We can characterize stocks along all of these dimensions. The linear program will then attempt to build portfolios that are reasonably close to the benchmark portfolio in all of the dimensions used for risk control.

It is also possible to set up a linear program with explicit transactions costs, a limit on turnover, and upper and lower position limits on each stock. The objective of the linear program is to maximize the portfolio's alpha less transactions costs, while remaining close to the benchmark portfolio in the risk control dimensions.

The linear program takes all the information about alpha into account and controls risk by keeping the characteristics of the portfolio close to the characteristics of the benchmark. However, the linear program has difficulty producing portfolios with a prespecified number of stocks. Also, the risk-control characteristics should not work at cross purposes with the alphas. For example, if the alphas tell you to shade the portfolio toward smaller stocks at some times and toward larger stocks at other times, you should not control risk on the size dimension.

Quadratic Programming

Quadratic programming (QP) is the ultimate[10] in portfolio construction. The quadratic program explicitly considers each of the three elements in our figure of merit: alpha, risk, and transactions costs. In addition, since a quadratic program includes a linear program as a special case, it can include all the constraints and limitations one finds in a linear program. This should be the best of all worlds. Alas, nothing is perfect.

One of the main themes of this chapter is dealing with less than perfect data. The quadratic program requires a great many more inputs than the other portfolio construction techniques. More inputs mean more noise. Does the benefit of explicitly considering risk outweigh the cost of introducing additional noise? A universe of 500 stocks will require 500 volatility estimates and 124,750 correlation estimates. There are ample opportunities to make mistakes. It is a fear of garbage in, garbage out that deters managers from using a quadratic program.

This fear is warranted. A lack of precision in the estimate of correlations is an inconvenience in the ordinary estimation of portfolio risk. For the most part, the estimation errors will cancel out. It is an obstacle in optimization. In optimization, the portfolio

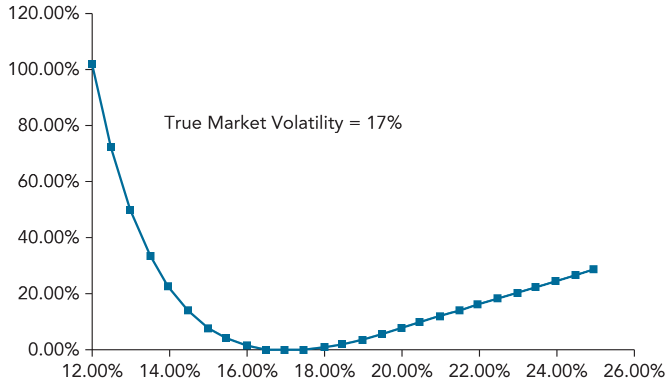


Figure 4.2 Estimated market volatility.

is selected to, among other things, have a low level of active risk. Because the optimizer tries to lower active risk, it will take advantage of opportunities that appear in the noisy estimates of covariance but are not present in reality.

An example can illustrate the point. Suppose we consider a simple cash versus market trade-off. Let ζ be the actual volatility of the market and σ our perceived volatility. If VA* is the optimal value added that we can obtain with the correct risk estimate ζ, then the loss we obtain with the estimate σ is

$$\text{Loss} = \text{VA}^* \cdot \left[1 - \left(\frac{\zeta}{\sigma}\right)^2\right]^2 \tag{4.11}$$

Figure 4.2 shows the percentage loss, Loss/VA*, as a function of the estimated market risk, assuming that the true market risk is 17 percent. In this example, market volatility estimates within 1 percent of the true market volatility will not hurt value added very much, but as estimation error begins to exceed 3 percent, the effect on value added becomes significant, especially if the error is an underestimate of volatility. In fact, an underestimate of 12 percent market volatility (5 percent below the "true" volatility) leads to a negative value added.

There are two lessons here. The first is that errors in the estimates of covariance lead to inefficient implementation. The second, which is more positive and, indeed, more important, is that it is vital to have good estimates of covariance. Rather than abandon the attempt, try to do a good job.

4.8 TESTS OF PORTFOLIO CONSTRUCTION METHODS

We can test the effectiveness of these portfolio construction procedures by putting them on an equal footing and judging the performance of their outputs. In this case, we will input

[10] Given our criterion of portfolio alpha minus a penalty for active risk and less transactions costs.

identical alphas to four procedures, described below, and ignore transactions costs.[11]

The alphas are great. They include the actual returns to the 500 stocks in the S&P 500 over the next year plus noise, combined so that the correlation of the alphas with the returns (the information coefficient) is 0.1. The fundamental law of active management therefore predicts[12] an information ratio of 2.24. So not only will we feed the same alphas into each portfolio construction method, but we know what the final result should be.

The four portfolio construction techniques are:

- *Screen I.* Take the N stocks with the highest alphas and equal-weight them. Use $N = 50$, 100, and 150 for low, medium, and high risk aversion, respectively.
- *Screen II.* Take the N stocks with the highest alphas and capitalization-weight them. Use $N = 50$, 100, and 150 for low, medium, and high risk aversion, respectively.
- *Strat.* Take the J stocks with the highest alphas in each of the BARRA 55 industry categories. Use $J = 1$, 2, and 3 for low, medium, and high risk aversion portfolios, which will have 55, 110, and 165 stocks, respectively.
- *QP.* Choose portfolios which maximize value added, assuming low, medium, and high risk aversion parameters. Use full investment and no short sales constraints, and constrain each position to constitute no more than 10 percent of the entire portfolio.

Portfolios were constructed in January 1984 and rebalanced in January 1985, January 1986, and May 1987, with portfolio performance tracked over the subsequent year. Table 4.3 contains the results.

Table 4.3 displays each portfolio's ex post information ratio. In this test, the quadratic programming approach clearly led to consistently the highest ex post information ratios. On average, it surpassed all the other techniques, and it exhibited consistent performance around that average. A stratified portfolio had the single highest ex post information ratio, but no consistency over time. The screening methods in general do not methodically control for risk, and Table 4.3 shows that one of the screened portfolios even experienced negative returns during one period.

Recall that the ex ante target for the information ratio was 2.24. None of the methods achieved that target, although the quadratic program came closest on average. Part of the reason for the shortfall is the constraints imposed on the optimizer. We calculated the target information ratio ignoring constraints. As we have seen, constraints can effectively reduce the information coefficient and hence the information ratio.

[11] For more details, see Muller (1993). We ignore transactions costs to simplify the test.

[12] The information coefficient of 0.1 and the breadth of 500 leads to $IR = 0.1 \cdot \sqrt{500} = 2.24$.

Table 4.3

Date	Risk Aversion	Screen I	Screen II	Strat	QP
January 1984	High	1.10	1.30	0.63	2.16
	Medium	0.95	2.24	0.64	1.89
	Low	0.73	1.31	0.69	1.75
January 1985	High	0.78	1.47	1.98	0.98
	Medium	0.74	−0.53	1.29	1.68
	Low	0.50	−0.15	0.83	1.49
January 1986	High	1.17	0.91	0.69	2.08
	Medium	0.69	0.98	0.33	2.29
	Low	0.60	0.99	0.51	2.51
May 1987	High	1.43	2.04	2.82	2.14
	Medium	1.01	1.48	2.60	1.76
	Low	0.66	1.17	2.17	1.82
Average		0.86	1.10	1.27	1.88
Standard deviation		0.27	0.79	0.89	0.40
Maximum		1.43	2.24	2.82	2.51
Minimum		0.50	−0.53	0.33	0.98

Source: Peter Muller, "Empirical Tests of Biases in Equity Portfolio Optimization," in *Financial Optimization*, edited by Stavros A. Zenios (Cambridge: Cambridge University Press, 1993), Table 4-4.

4.9 ALTERNATIVES TO MEAN/VARIANCE OPTIMIZATION

Alternatives to standard deviation as risk measurements include semivariance, downside risk, and shortfall probability. We reviewed the alternatives and chose standard deviation as the best overall risk measure. We return to the issue again here, since our portfolio construction objective expresses our utility, which may in fact depend on alternative measures of risk. But as two research efforts show, even if your personal preferences depend on alternative risk measures, mean/variance analysis will produce equivalent or better portfolios. We present the research conclusions here, and cite the works in the bibliography.

Kahn and Stefek (1996) focus on the forward-looking nature of portfolio construction. The utility function includes forecasts

of future risk. Mean/variance analysis, as typically applied in asset selection, relies on sophisticated modeling techniques to accurately forecast risk.

Forecasting of alternative risk measures must rely on historical returns–based analysis. Kahn and Stefek show that higher moments of asset and asset class return distributions exhibit very little predictability, especially where it is important for portfolio construction. Return kurtosis is predictable, in the sense that most return distributions exhibit positive kurtosis ("fat tails") most of the time. The ranking of assets or asset classes by kurtosis exhibits very little predictability. The only exception is options, where return asymmetries are engineered into the payoff pattern.

The empirical result is that most alternative risk forecasts reduce to a standard deviation forecast plus noise, with even the standard deviation forecast based only on history. According to this research, even investors with preferences defined by alternative risk measures are better served by mean/variance analysis.[13]

Grinold (1999) takes a different approach to the same problem, in the specific case of asset allocation. First, he adjusts returns-based analysis to the institutional context: benchmark-aware investing with typical portfolios close to the benchmark. This is the same approach we have applied to mean/variance analysis in this text. Then he compares mean/variance and returns-based analysis, assuming that the benchmark holds no options and that all options are fairly priced.

The result is that portfolios constructed using returns-based analysis are very close to mean/variance portfolios, although they require much more effort to construct. Furthermore, managers using this approach very seldom buy options. If options are fairly priced relative to the underlying asset class, then optimization will pursue the alphas directly through the asset class, not indirectly through the options.

So Kahn and Stefek argue the asset selection case for mean/variance, and Grinold argues the asset allocation case for mean/variance. Furthermore, Grinold shows why institutional investors, with their aversion to benchmark risk, will seldom purchase options—the only type of asset requiring analysis beyond mean/variance.

As a final observation, though, some active institutional investors do buy options. We argue that they do so typically to evade restrictions on leverage or short selling, or because of liquidity concerns. Only in the case of currency options do we see much evidence of investors choosing options explicitly for their distributions. Many managers have a great aversion to currency losses, and options can provide downside protection. We still advocate using mean/variance analysis generally and, if necessary, treating currency options as a special case.

4.10 DISPERSION

Dispersion plagues every manager running separate accounts for multiple clients. Each account sees the same alphas, benchmark, and investment process. The cash flows and history differ, however, and the portfolios are not identical. Hence, portfolio returns are not identical.

We will define *dispersion* as the difference between the maximum return and minimum return for these separate account portfolios. If the holdings in each account are identical, dispersion will disappear. If transactions costs were zero, dispersion would disappear. Dispersion is a measure of how an individual client's portfolio may differ from the manager 's reported composite returns. Dispersion is, at the least, a client support problem for investment managers.

In practice, dispersion can be enormous. We once observed five investors in a particular manager 's strategy, in separate accounts, incur dispersion of 23 percent over a year. The manager 's overall dispersion may have been even larger. This was just the dispersion involving these five clients. In another case, with another manager, one client outperformed the S&P 500 by 15 percent while another underperformed by 9 percent, in the same year. At that level, dispersion is much more than a client support problem.

We can classify dispersion by its various sources. The first type of dispersion is client-driven. Portfolios differ because individual clients impose different constraints. One pension fund may restrict investment in its company stock. Another may not allow the use of futures contracts. These client-initiated constraints lead to dispersion, but they are completely beyond the manager's control.

But managers can control other forms of dispersion. Often, dispersion arises through a lack of attention. Separate accounts exhibit different betas and different factor exposures through lack of attention. Managers should control this form of dispersion.

On the other hand, separate accounts with the same factor exposures and betas can still exhibit dispersion because of owning different assets. Often the cost of holding exactly the same assets in each account will exceed any benefit from reducing dispersion.

In fact, because of transactions costs, some dispersion is optimal. If transactions costs were zero, rebalancing all the separate

[13] The case of investors in options and dynamic strategies like portfolio insurance is a bit trickier, but also handled in the paper. There the conclusion is to apply mean/variance analysis to the active asset selection strategy, and to overlay an options-based strategy based on alternative risk measures. But see Grinold (1999), who shows that under reasonable assumptions, even with alternative risk measures, most institutional investors will not use such strategies.

accounts so that they hold exactly the same assets in the same proportions would have no cost. Dispersion would disappear, at no cost to investors. With transactions costs, however, managers can achieve zero dispersion only with increased transactions costs. Managers should reduce dispersion only until further reduction would substantially lower returns on average because much higher transactions costs would be incurred.

Example

To understand dispersion better, let's look at a concrete example. In this example, the manager runs an existing portfolio and receives cash to form a new portfolio investing in the same strategy. So at one point in time, the manager is both rebalancing the existing portfolio and constructing the new portfolio. The rebalanced portfolio holdings will reflect both new and old information. With zero transactions costs, the manager would rebalance to the new optimum. Given an existing portfolio, though, he rebalances only where the new information more than overcomes the transactions costs, as in Equation (4.9).

This trade-off does not affect the new portfolio in the same way. The manager starts from cash, and while he would still like to minimize transactions costs, he assumes a fairly high transactions cost for the initial portfolio construction. For this example, we'll assume that the new portfolio he builds is optimal and reflects entirely the manager 's new information.

Clearly there will be dispersion between the existing portfolio and the new portfolio. There are two methods by which the manager could reduce dispersion to zero. He could invest the new portfolio in the rebalanced existing portfolio. This sacrifices returns, since the new portfolio will reflect both new and old information instead of just new information. The other choice is to invest the composite in the new optimum. But this would require paying excess transactions costs. By treating the existing portfolio and the new portfolio separately, the manager accepts some level of dispersion in order to achieve higher average returns. Furthermore, he can hope that this dispersion will decrease over time.

Characterizing Dispersion

We will now perform some static analysis to understand the causes of dispersion. First, consider dispersion caused by different betas or factor exposures. If the separate account betas range from 0.9 to 1.1 and the market return is 35 percent one year, then the dispersion would be 7 percent based just on the differing betas. This range of betas is quite large for an efficient, quantitatively run optimal process, and yet it doesn't come close to explaining some of the extreme war stories.

Now let's consider static analysis of managed dispersion—where the manager has matched factor exposures but not assets across all accounts—to try to understand the magnitude of the effect. In this simple model, we will consider N portfolios, all equally weighted with identical factor exposures. Each portfolio contains 100 stocks, and out of that 100 stocks, M stocks appear in all the portfolios and $100 - M$ stocks are unique to the particular portfolio. Furthermore, every stock has identical specific risk of 20 percent. Figure 4.3 displays the results, assuming normal distributions.

We can use the model to show that dispersion will depend on the number of stocks the portfolios have in common, the overall levels of specific risk, and the overall number of portfolios under management.

Managing Dispersion

We have seen how some level of dispersion is optimal and have discussed why dispersion arises. The next question is whether dispersion decreases over time: Do dispersed portfolios converge, and how fast? In general, convergence will depend on the type of alphas in the strategy, the transactions costs, and possibly the portfolio construction methodology.

If alphas and risk stay absolutely constant over time, then dispersion will never disappear. There will always be a transactions cost barrier. An exact matching of portfolios will never pay. Furthermore, we can show that the remaining

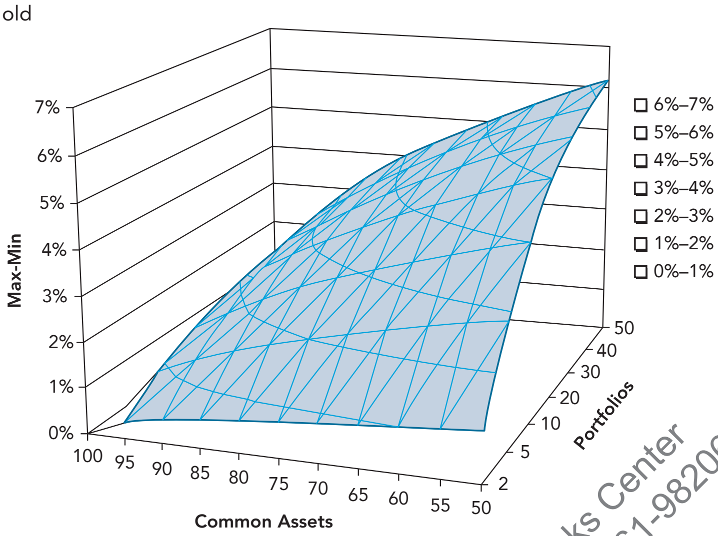


Figure 4.3 Dispersion: 100 stock portfolios.

tracking error is bounded based on the transactions costs and the manager's risk aversion:

$$\psi^2 \leq \frac{TC}{2 \cdot \lambda_A} \tag{4.12}$$

where TC measures the cost of trading from the initial portfolio to the zero transactions cost optimal portfolio (which we will refer to as portfolio Q), and we are measuring tracking error *and* risk aversion relative to portfolio Q. With very high risk aversion, all portfolios must be close to one another. But the higher the transactions costs, the more tracking error there is. Given intermediate risk aversion of $\lambda_A = 0.10$ and round-trip transactions costs of 2 percent, and assuming that moving from the initial portfolio to portfolio Q involves 10 percent turnover, Equation (4.12) implies tracking error of 1 percent.

Since tracking error is bounded, dispersion is also bounded. Dispersion is proportional to tracking error, with the constant of proportionality dependent on the number of portfolios being managed:

$$E\{r_{PA,\max} - r_{PA,\min}\} = 2 \cdot \Phi^{-1}\left\{\left(\frac{1}{2}\right)^{1/N}\right\} \cdot \psi \tag{4.13}$$

where this constant of proportionality involves the inverse of the cumulative normal distribution function Φ, and ψ is the tracking error of each portfolio relative to the composite. Figure 4.3 displays this function. For a given tracking error, more portfolios lead to more dispersion because more portfolios will further probe the extremes of the return distribution.

If the alphas and risk vary over time—the usual case—then convergence will occur. We can show that with changing alphas and risk each period, the portfolios will either maintain or, more typically, decrease the amount of dispersion. Over time, the process inexorably leads to convergence, because each separate account portfolio is chasing the same moving target. These general arguments do not, however, imply any particular time scale.

As an empirical example, we looked at five U.S. equity portfolios in a strategy with alphas based on book-to-price ratios and stock-specific alphas. Roughly two-thirds of the strategy's value came from the book-to-price factor tilt, with one-third arising from the stock-specific alphas. We started these five portfolios in January 1992 with 100 names in each portfolio, but not the same 100 names in each portfolio. Each portfolio had roughly a 3 percent tracking error relative to the S&P 500. We analyzed the initial level of dispersion and then looked at how that changed over time. We used a consistent alpha generation process and standard mean/variance optimization with uniform transactions costs. To understand convergence and transactions costs, we looked at behavior as we changed the overall level of transactions costs.

What we found was a steady decrease in average tracking error (relative to the composite) and dispersion, with the smallest dispersion exhibited when we assumed the lowest transactions costs. Figure 4.4 displays the results. So even though

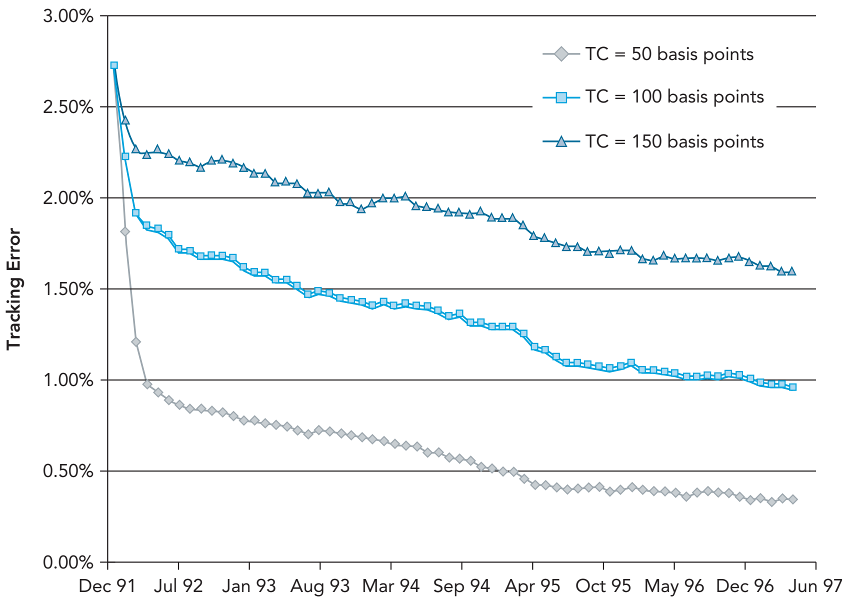


Figure 4.4 Convergence. (*Source:* BARRA.)

our starting portfolios differed, they steadily converged over a roughly 5-year period. In real-life situations, client-initiated constraints and client-specific cash flows will act to keep separate accounts from converging.

One final question is whether we can increase convergence by changing our portfolio construction technology. In particular, what if we used dual-benchmark optimization? Instead of penalizing only active risk relative to the benchmark, we would also penalize active risk relative to either the composite portfolio or the optimum calculated ignoring transactions costs.

Dual-benchmark optimization can clearly reduce dispersion, but only at an undesirable price. Dual-benchmark optimization simply introduces the trade-off we analyzed earlier: dispersion versus return. Unless you are willing to give up return in order to lower dispersion, do not implement the dual-benchmark optimization approach to managing dispersion.

SUMMARY

The theme of this chapter has been portfolio construction in a less than perfect world. We have taken the goals of the portfolio manager as given. The manager wants the highest possible after-cost value added. The before-cost value added is the portfolio's alpha less a penalty for active variance. The costs are for the transactions needed to maintain the portfolio's alpha.

Understanding and achieving this goal requires data on alphas, covariances between stock returns, and estimates of transactions costs. Alpha inputs are often unrealistic and biased. Covariances and transactions costs are measured imperfectly.

In this less than perfect environment, the standard reaction is to compensate for flawed inputs by regulating the outputs of the portfolio construction process: placing limits on active stock positions, limiting turnover, and constraining holdings in certain categories of stocks to match the benchmark holdings.

These are valid approaches, as long as we recognize that their purpose is to compensate for faulty inputs. We prefer a direct attack on the causes. Treat flaws in the alpha inputs with alpha analysis: Remove biases, trim outlandish values, and scale alphas in line with expectations for value added. This strengthens the link between research and portfolio construction. Then seek out the best possible estimates of risk and transactions costs. As appropriate, use a powerful portfolio construction tool with as few added constraints as possible.

Near the end of the chapter, we returned to the topic of alternative risk measures and alternatives to mean/variance optimization. For most active institutional managers (especially those who do not invest in options and optionlike dynamic strategies such as portfolio insurance), alternatives to mean/variance analysis greatly complicate portfolio construction without improving results. At the stock selection level, results may be much worse.

Finally, we analyzed the very practical issue of dispersion among separately managed accounts. We saw that managers can control dispersion—especially that driven by differing factor exposures—but should not reduce it to zero.

References

Chopra, Vijay K., and William T. Ziemba. "The Effects of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." *Journal of Portfolio Management,* vol. 19, no. 2, 1993, pp. 6–11.

Connor, Gregory, and Hayne Leland. "Cash Management for Index Tracking." *Financial Analysts Journal,* vol. 51, no. 6, 1995, pp. 75–80.

Grinold, Richard C. "The Information Horizon." *Journal of Portfolio Management,* vol. 24, no. 1, 1997, pp. 57–67.

———. "Mean-Variance and Scenario-Based Approaches to Portfolio Selection." *Journal of Portfolio Management,* vol. 25, no. 2, Winter 1999, pp. 10–22.

Jorion, Philippe. "Portfolio Optimization in Practice." *Financial Analysts Journal,* vol. 48, no. 1, 1992, pp. 68–74.

Kahn, Ronald N. "Managing Dispersion." BARRA Equity Research Seminar, Pebble Beach, Calif., June 1997.

Kahn, Ronald N., and Daniel Stefek. "Heat, Light, and Downside Risk." *BARRA Preprint,* December 1996.

Leland, Hayne. *Optimal Asset Rebalancing in the Presence of Transactions Costs.* University of California Research Program in Finance, Publication 261, October 1996.

Michaud, Richard. "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal,* vol. 45, no. 1, 1989, pp. 31–42.

Muller, Peter. "Empirical Tests of Biases in Equity Portfolio Optimization." In *Financial Optimization,* edited by Stavros A. Zenios (Cambridge: Cambridge University Press, 1993), pp. 80–98.

Rohweder, Herold C. "Implementing Stock Selection Ideas: Does Tracking Error Optimization Do Any Good?" *Journal of Portfolio Management,* vol. 24, no. 3, 1998, pp. 49–59.

Rudd, Andrew. "Optimal Selection of Passive Portfolios." *Financial Management,* vol. 9, no. 1, 1980, pp. 57–66.

Rudd, Andrew, and Barr Rosenberg. "Realistic Portfolio Optimization." *TIMS Study in the Management Sciences,* vol. 11, 1979, pp. 21–46.

Stevens, Guy V. G. "On the Inverse of the Covariance Matrix in Portfolio Analysis." *Journal of Finance,* vol. 53, no. 5, 1998, pp. 1821–1827.

5

Portfolio Risk: Analytical Methods

Learning Objectives

After completing this reading you should be able to:

- Define, calculate, and distinguish between the following portfolio VaR measures: diversified and undiversified portfolio VaR, individual VaR, incremental VaR, marginal VaR, and component VaR.
- Explain the impact of correlation on portfolio risk.
- Apply the concept of marginal VaR in making portfolio management decisions.
- Explain the risk-minimizing position and the risk and return-optimizing position of a portfolio.
- Explain the difference between risk management and portfolio management and describe how to use marginal VaR in portfolio management.

Excerpt is Chapter 7 of Value at Risk: The New Benchmark for Managing Financial Risk, *Third Edition, by Philippe Jorion.*

> Trust not all your goods to one ship.
>
> —Erasmus

Absent any insight into the future, prudent investors should diversify across sources of financial risk. This was the message of portfolio analysis laid out by Harry Markowitz in 1952. Thus the concept of value-at-risk (VaR), or portfolio risk, is not new. What is new is the systematic application of VaR to many sources of financial risk, or portfolio risk. VaR explicitly accounts for leverage and portfolio diversification and provides a simple, single measure of risk based on current positions.

There are many approaches to measuring VaR. The shortest road assumes that asset payoffs are linear (or delta) functions of normally distributed risk factors. Indeed, the *delta-normal method* is a direct application of traditional portfolio analysis based on variances and covariances, which is why it is sometimes called the *covariance matrix approach*.

This approach is *analytical* because VaR is derived from closed-form solutions. The analytical method developed in this chapter is very useful because it creates a more intuitive understanding of the drivers of risk within a portfolio. It also lends itself to a simple decomposition of the portfolio VaR.

This chapter shows how to measure and manage portfolio VaR. The first section details the construction of VaR using information on positions and the covariance matrix of its constituent components.

The fact that portfolio risk is not cumulative provides great diversification benefits. To manage risk, however, we also need to understand what will reduce it. The section that follows provides a detailed analysis of VaR tools that are essential to control portfolio risk. These include marginal VaR, incremental VaR, and component VaR. These VaR tools allow users to identify the asset that contributes most to their total risk, to pick the best hedge, to rank trades, or in general, to select the asset that provides the best risk-return trade-off. Then, a fully worked out example of VaR computations for a global equity portfolio and for Barings' fatal positions will be presented.

The advantage of analytical models is that they provide closed-form solutions that help our intuition. The methods presented here, however, are quite general. We will show how to build these VaR tools in a nonparametric environment. This applies to simulations, for example.

Finally, we will be taken toward portfolio optimization, which should be the ultimate purpose of VaR. We first show how the passive measurement of risk can be extended to the management of risk, in particular, risk minimization. We then integrate risk with expected returns and show how VaR tools can be use to move the portfolio toward the best combination of risk and return.

5.1 PORTFOLIO VAR

A portfolio can be characterized by positions on a certain number of constituent assets, expressed in the base currency, say, dollars. If the positions are fixed over the selected horizon, the portfolio rate of return is a *linear* combination of the returns on underlying assets, where the weights are given by the relative amounts invested at the beginning of the period. Therefore, the VaR of a portfolio can be constructed from a combination of the risks of underlying securities.

Define the portfolio rate of return from t to $t+1$ as

$$R_{p,t+1} = \sum_{i=1}^{N} w_i R_{i,t+1} \tag{5.1}$$

where N is the number of assets, $R_{i,t+1}$ is the rate of return on asset i, and w_i is the weight. The *rate of return* is defined as the change in the dollar value, or dollar return, scaled by the initial investment. This is a unitless measure.

Weights are constructed to sum to unity by scaling the dollar positions in each asset W_i by the portfolio total market value W. This immediately rules out portfolios that have zero net investment $W = 0$, such as some derivatives positions. But we could have positive and negative weights w_i, including values much larger than 1, as with a highly leveraged hedge fund. If the net portfolio value is zero, we could use another measure, such as the sum of the gross positions or absolute value of all dollar positions W^*. All weights then would be defined in relation to this benchmark. Alternatively, we could express returns in dollar terms, defining a dollar amount invested in asset i as $W_i = w_i W$. We will be using x as representing the vector of dollar amount invested in each asset so as to avoid confusion with the total dollar amount W.

It is important to note that in traditional mean-variance analysis, each constituent asset is a security. In contrast, VaR defines the component as a *risk factor* and w_i as the linear exposure to this risk factor. Whether dealing with assets or risk factors, the mathematics of portfolio VaR are equivalent, however.

To shorten notation, the portfolio return can be written using *matrix notation*, replacing a string of numbers by a single vector:

$$R_p = w_1R_1 + w_2R_2 + \cdots + w_NR_N = [w_1 w_2 \cdots w_N] \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix} = w'R \tag{5.2}$$

where w' represents the transposed vector (i.e., horizontal) of weights, and R is the vertical vector containing individual asset returns.

The portfolio expected return is

$$E(R_p) = \mu_p = \sum_{i=1}^{N} w_i \mu_i \tag{5.3}$$

and the variance is

$$V(R_p) = \sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{j=1, j \neq 1}^{N} w_i w_j \sigma_{ij}$$

$$= \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j<1}^{N} w_i w_j \sigma_{ij} \tag{5.4}$$

This sum accounts not only for the risk of the individual securities σ_i^2 but also for all covariances, which add up to a total of $N(N-1)/2$ different terms.

As the number of assets increases, it becomes difficult to keep track of all covariance terms, which is why it is more convenient to use matrix notation. The variance can be written as

$$\sigma_p^2 = [w_1 \cdots w_N] \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1N} \\ \vdots & & & & \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \cdots & \sigma_{N2} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}$$

Defining Σ as the covariance matrix, the variance of the portfolio rate of return can be written more compactly as

$$s_p^2 = w'\Sigma w \tag{5.5}$$

where w are weights, which have no units. This also can be written in terms of dollar exposures x as

$$s_p^2 W^2 = x'\Sigma x \tag{5.6}$$

So far nothing has been said about the distribution of the portfolio return. Ultimately, we would like to translate the portfolio variance into a VaR measure. To do so, we need to know the distribution of the portfolio return. In the delta-normal model, all individual security returns are assumed normally distributed. This is particularly convenient because the portfolio return, a linear combination of jointly normal random variables, is also normally distributed. If so, we can translate the confidence level c into a standard normal deviate α such that the probability of observing a loss worse than $-\alpha$ is c. Defining W as the initial portfolio value, the portfolio VaR is

$$\text{Portfolio VaR} = \text{VaR}_p = \alpha \sigma_p W = \alpha \sqrt{x'\Sigma x} \tag{5.7}$$

Diversified VaR The portfolio VaR, taking into account diversification benefits between components.

At this point, we also can define the individual risk of each component as

$$\text{VaR}_i = \alpha \sigma_i |W_i| = \alpha \sigma_i |w_i| W \tag{5.8}$$

Note that we took the absolute value of the weight w_i because it can be negative, whereas the risk measure must be positive.

Individual VaR The VaR of one component taken in isolation.

Equation (5.4) shows that the portfolio VaR depends on variances, covariances, and the number of assets. Covariance is a measure of the extent to which two variables move linearly together. If two variables are independent, their covariance is equal to zero. A positive covariance means that the two variables tend to move in the same direction; a negative covariance means that they tend to move in opposite directions. The magnitude of covariance, however, depends on the variances of the individual components and is not easily interpreted. The *correlation coefficient* is a more convenient, scale-free measure of linear dependence:

$$\rho_{12} = \sigma_{12}/(\sigma_1 \sigma_2) \tag{5.9}$$

The correlation coefficient ρ always lies between -1 and $+1$. When equal to unity, the two variables are said to be *perfectly correlated*. When 0, the variables are *uncorrelated*.

Lower portfolio risk can be achieved through low correlations or a large number of assets. To see the effect of N, assume that all assets have the same risk and that all correlations are the same, that equal weight is put on each asset. Figure 5.1 shows how portfolio risk decreases with the number of assets.

Start with the risk of one security, which is assumed to be 20 percent. When ρ is equal to zero, the risk of a 10-asset portfolio drops to 6.3 percent; increasing N to 100 drops the risk even further to 2.0 percent. Risk tends asymptotically to zero. More generally, portfolio risk is

$$\sigma_p = \sigma \sqrt{\frac{1}{N} + \left(1 - \frac{1}{N}\right)\rho} \tag{5.10}$$

which tends to $\sigma\sqrt{\rho}$ as N increases. Thus, when $\rho = 0.5$, risk decreases rapidly from 20 to 14.8 percent as N goes to 10 and afterward converges more slowly toward its minimum value of 14.1 percent.

Low correlations thus help to diversify portfolio risk. Take a simple example with two assets only. The "diversified" portfolio variance is

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2 \tag{5.11}$$

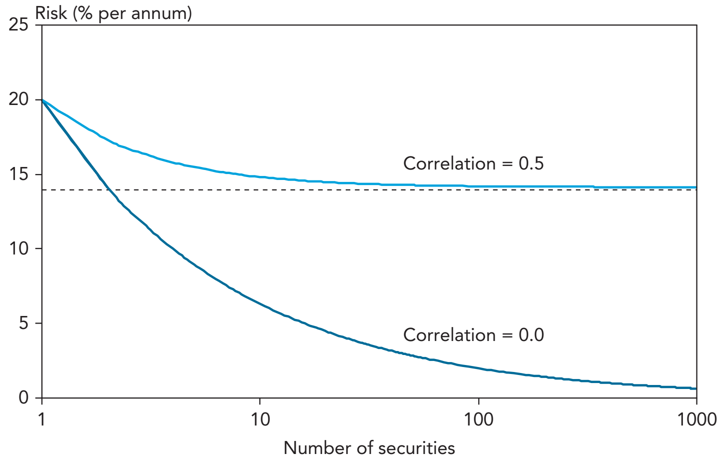


Figure 5.1 Risk and number of securities.

The portfolio VaR is then

$$\mathrm{VaR}_p = \alpha\sigma_p W = \alpha\sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}W \quad (5.12)$$

This can be related to the individual VaR as defined in Equation (5.8).

When the correlation ρ is zero, the portfolio VaR reduces to

$$\mathrm{VaR}_p = \sqrt{\alpha^2 w_1^2 W^2\sigma_1^2 + \alpha^2 w_2^2 W^2 \sigma_2^2} = \sqrt{\mathrm{VaR}_1^2 + \mathrm{VaR}_2^2} \quad (5.13)$$

The portfolio risk must be lower than the sum of the individual VaRs: $\mathrm{VaR}_p < \mathrm{VaR}_1 + \mathrm{VaR}_2$. This reflects the fact that with assets that move independently, a portfolio will be less risky than either asset. Thus VaR is a *coherent* risk measure for normal and, more generally, elliptical distributions.

When the correlation is exactly unity and w_1 and w_2 are both positive, Equation (5.12) reduces to

$$\mathrm{VaR}_p = \sqrt{\mathrm{VaR}_1^2 + \mathrm{VaR}_2^2 + 2\mathrm{VaR}_1 \times \mathrm{VaR}_2}$$
$$= \mathrm{VaR}_1 + \mathrm{VaR}_2 \quad (5.14)$$

In other words, the portfolio VaR is equal to the sum of the individual VaR measures if the two assets are perfectly correlated. In general, though, this will not be the case because correlations typically are imperfect. The benefit from diversification can be measured by the difference between the *diversified* VaR and the *undiversified* VaR, which typically is shown in VaR reporting systems.

Undiversified VaR The sum of individual VaRs, or the portfolio VaR when there is no short position and all correlations are unity.

This interpretation differs when short sales are allowed. Suppose that the portfolio is long asset 1 but short asset 2 (w_1 is positive, and w_2 is negative). This could represent a hedge fund that has \$1 in capital and a \$1 billion long position in corporate bonds and a \$1 billion short position in Treasury bonds, the rationale for the position being that corporate yields are slightly higher than Treasury yields. If the correlation is exactly unity, the fund has no risk because any loss in one asset will be offset by a matching gain in the other. The portfolio VaR then is zero.

Instead, the risk will be greatest if the correlation is −1, in which case losses in one asset will be amplified by the other. Here, the *undiversified* VaR can be interpreted as the portfolio VaR when the correlation attains its worst value, which is −1. Therefore, the undiversified VaR provides an upper bound on the portfolio VaR should correlations prove unstable and all move at the same time in the wrong direction. It provides an absolute worst-case scenario for the portfolio at hand.

Example 5.1

Consider a portfolio with two foreign currencies, the Canadian dollar (CAD) and the euro (EUR). Assume that these two currencies are uncorrelated and have a volatility against the dollar of 5 and 12 percent, respectively. The first step is to mark to market the positions in the base currency. The portfolio has US\$2 million invested in the CAD and US\$1 million in the EUR. We seek to find the portfolio VaR at the 95 percent confidence level.

First, we will compute the variance of the portfolio dollar return. Define x as the dollar amounts allocated to each risk factor, in millions. Compute the product

$$\Sigma x = \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.12 \end{bmatrix}\begin{bmatrix} \$2 \\ \$1 \end{bmatrix} = \begin{bmatrix} 0.05^2 \times \$2 + 0 \times \$1 \\ 0 \times \$2 + 0.12^2 \times \$1 \end{bmatrix}$$
$$= \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix}$$

The portfolio variance then is (in dollar units)

$$\sigma_p^2 W^2 = x'(\Sigma x) = [\$2\ \$1]\begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} = 0.0100 + 0.0144 = 0.0244$$

The dollar volatility is $\sqrt{0.0244} = \$0.156205$ million. Using $\alpha = 1.65$, we find $\mathrm{VaR}_p = 1.65 \times 156{,}205 = \$257{,}738$.

Next, the individual (undiversified) VaR is found simply as $\mathrm{VaR}_i = \alpha\sigma_i x_i$, that is,

$$\begin{bmatrix} \mathrm{VaR}_1 \\ \mathrm{VaR}_2 \end{bmatrix} = \begin{bmatrix} 1.65 \times 0.05 \times \$2 \text{ million} \\ 1.65 \times 0.12 \times \$1 \text{ million} \end{bmatrix} = \begin{bmatrix} \$165{,}000 \\ \$198{,}000 \end{bmatrix}$$

Note that these numbers sum to an undiversified VaR of \$363,000, which is greater than the portfolio VaR of \$257,738 owing to diversification effects.

5.2 VaR TOOLS

Initially, VaR was developed as a methodology to measure portfolio risk. There is much more to VaR than simply reporting a single number, however. Over time, risk managers have discovered that they could use the VaR process for active risk management. A typical question may be, "Which position should I alter to modify my VaR most effectively?" Such information is quite useful because portfolios typically are traded incrementally owing to transaction costs. This is the purpose of VaR tools, which include marginal, incremental, and component VaR.

Marginal VaR

To measure the effect of changing positions on portfolio risk, individual VaRs are not sufficient. Volatility measures the uncertainty in the return of an asset, taken in isolation. When this asset belongs to a portfolio, however, what matters is the contribution to portfolio risk.

We start from the existing portfolio, which is made up of N securities, numbered as $j = 1, \ldots, N$. A new portfolio is obtained by adding one unit of security i. To assess the impact of this trade, we measure its "marginal" contribution to risk by increasing w by a small amount or differentiating Equation (5.4) with respect to w_i, that is,

$$\frac{\partial \sigma_p^2}{\partial w_i} = 2w_i\sigma_i^2 + 2\sum_{j=1, j\neq i}^{N} w_j\sigma_{ij}$$

$$= 2\mathrm{cov}(R_i, w_iR_i + \sum_{j\neq i}^{N} w_jR_j) = 2\mathrm{cov}(R_i, R_p) \qquad (5.15)$$

Instead of the derivative of the variance, we need that of the volatility. Noting that $\partial\sigma_p^2/\partial w_i = 2\sigma_p\partial\sigma_p/\partial w_i$, the sensitivity of the portfolio volatility to a change in the weight is then

$$\frac{\partial \sigma_p}{\partial w_i} = \frac{\mathrm{cov}(R_i, R_p)}{\sigma_p} \qquad (5.16)$$

Converting into a VaR number, we find an expression for the *marginal VaR*, which is a vector with component

$$\Delta \mathrm{VaR}_i = \frac{\partial \mathrm{VaR}}{\partial x_i} = \frac{\partial \mathrm{VaR}}{\partial w_i W} = \alpha\frac{\partial \sigma_p}{\partial w_i} = \alpha\frac{\mathrm{cov}(R_i, R_p)}{\sigma_p} \qquad (5.17)$$

Since this was defined as a ratio of the dollar amounts, this marginal VaR measure is unitless.

> **Marginal VaR** The change in portfolio VaR resulting from taking an additional dollar of exposure to a given component. It is also the partial (or linear) derivative with respect to the component position.

This marginal VaR is closely related to the *beta*, defined as

$$\beta_i = \frac{\mathrm{cov}(R_i, R_p)}{\sigma_p^2} = \frac{\sigma_{ip}}{\sigma_p^2} = \frac{\rho_{ip}\sigma_i\sigma_p}{\sigma_p^2} = \rho_{ip}\frac{\sigma_i}{\sigma_p} \qquad (5.18)$$

which measures the contribution of one security to total portfolio risk. Beta is also called the *systematic risk* of security i vis-à-vis portfolio p and can be measured from the slope coefficient in a regression of R_i on R_p, that is,

$$R_{i,t} = \alpha_i + \beta_i R_{p,t} + \epsilon_{i,t} \qquad t = 1, \ldots, T \qquad (5.19)$$

Using matrix notation, we can write the vector β, including all assets, as

$$\beta = \frac{\Sigma w}{(w'\Sigma w)}$$

Note that we already computed the vector Σw as an intermediate step in the calculation of VaR. Therefore, β and the marginal VaR can be derived easily once VaR has been calculated.

Beta risk is the basis for capital asset pricing model (CAPM) developed by Sharpe (1964). According to the CAPM, well-diversified investors only need to be compensated for the systematic risk of securities relative to the market. In other words, the risk premium on all assets should depend on beta only. Whether this is an appropriate description of capital markets has been the subject of much of finance research in the last decades. Even though this proposition is still debated hotly, the fact remains that systematic risk is a useful statistical measure of marginal portfolio risk.

To summarize, the relationship between the ΔVaR and β is

$$\Delta \mathrm{VaR}_i = \frac{\partial \mathrm{VaR}}{\partial x_i} = \alpha(\beta_i \times \sigma_p) = \frac{\mathrm{VaR}}{W} \times \beta_i \qquad (5.20)$$

The marginal VaR can be used for a variety of risk management purposes. Suppose that an investor wants to lower the portfolio VaR and has the choice to reduce all positions by a fixed amount, say, \$100,000. The investor should rank all marginal VaR numbers and pick the asset with the largest ΔVaR because it will have the greatest hedging effect.

Incremental VaR

This methodology can be extended to evaluate the total impact of a proposed trade on portfolio p. The new trade is represented by position a, which is a vector of additional exposures to our risk factors, measured in dollars.

Ideally, we should measure the portfolio VaR at the initial position VaR_p and then again at the new position VaR_{p+a}. The incremental VaR then is obtained, as described in Figure 5.2, as

$$\text{Incremental VaR} = \mathrm{VaR}_{p+a} - \mathrm{VaR}_p \qquad (5.21)$$

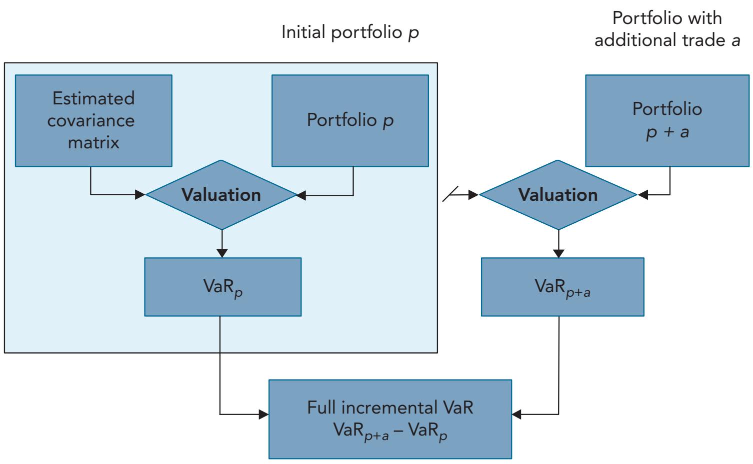


Figure 5.2 **The impact of a proposed trade with full revaluation.**

This "before and after" comparison is quite informative. If VaR is decreased, the new trade is risk-reducing or is a hedge; otherwise, the new trade is risk-increasing. Note that *a* may represent a change in a single component or a more complex trade with changes in multiple components. Hence, in general, *a* represents a vector of new positions.

Incremental VaR The change in VaR owing to a new position. It differs from the marginal VaR in that the amount added or subtracted can be large, in which case VaR changes in a nonlinear fashion.

The main drawback of this approach is that it requires a full revaluation of the portfolio VaR with the new trade. This can be quite time-consuming for large portfolios. Suppose, for instance, that an institution has 100,000 trades on its books and that it takes 10 minutes to do a VaR calculation. The bank has measured its VaR at some point during the day.

Then a client comes with a proposed trade. Evaluating the effect of this trade on the bank's portfolio again would require 10 minutes using the incremental-VaR approach. Most likely, this will be too long to wait to take action. If we are willing to accept an approximation, however, we can take a shortcut.[1]

Expanding VaR_{p+a} in series around the original point,

$$\text{VaR}_{p+a} = \text{VaR}_p + (\Delta\text{VaR})' \times a + \cdots \qquad (5.22)$$

where we ignored second-order terms if the deviations *a* are small. Hence the incremental VaR can be reported as, approximately,

$$\text{Incremental VaR} \approx (\Delta\text{VaR})' \times a \qquad (5.23)$$

This measure is much faster to implement because the ΔVaR vector is a by-product of the initial VaR_p computation. The new process is described in Figure 5.3.

Here we are trading off faster computation time against accuracy. How much of an improvement is this shortcut relative to the full incremental VaR method? The shortcut will be especially useful for large portfolios where a full revaluation requires a large number of computations. Indeed, the number of operations increases with the square of the number of risk factors. In addition, the shortcut will prove to be a good approximation for large portfolios where a proposed trade is likely to be small relative to the outstanding portfolio. Thus the simplified VaR method allows real-time trading limits.

The incremental VaR method applies to the general case where a trade involves a set of new exposures on the risk factors. Consider instead the particular case where a new trade involves a position in one risk factor only (or asset). The portfolio value changes from the old value of W to the new value of $W_{p+a} = W + a$, where a is the amount invested in asset i. We can write the variance of the dollar returns on the new portfolio as

$$\sigma_{p+a}^2 W_{p+a}^2 = \sigma_p^2 W^2 + 2aW\sigma_{ip} + a^2\sigma_i^2 \qquad (5.24)$$

An interesting question for portfolio managers is to find the size of the new trade that leads to the lowest portfolio risk. Differentiating with respect to a,

$$\frac{\partial \sigma_{p+a}^2 W_{p+a}^2}{\partial a} = 2W\sigma_{ip+2a\sigma_i^2} \qquad (5.25)$$

which attains a zero value for

$$a^* = -W\frac{\sigma_{ip}}{\sigma_i^2} = -W\beta_i\frac{\sigma_p^2}{\sigma_i^2} \qquad (5.26)$$

This is the variance-minimizing position, also known as best hedge.

Best hedge Additional amount to invest in an asset so as to minimize the risk of the total portfolio.

[1] See also Garman (1996 and 1997).

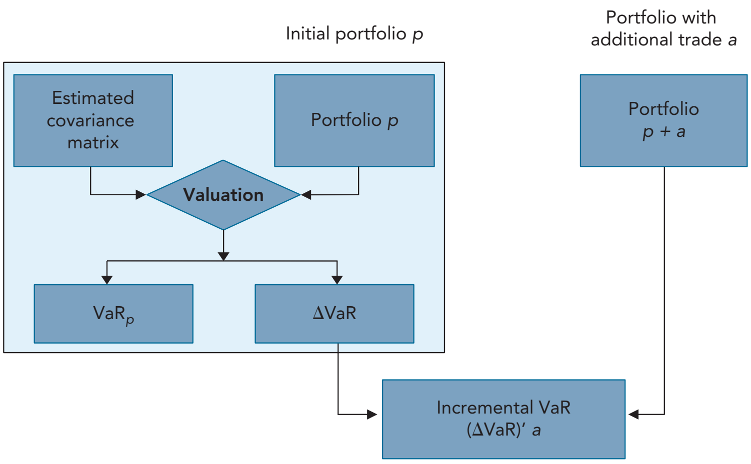


Figure 5.3 The impact of a proposed trade with marginal VaR.

Example 5.1 (continued)

Going back to the previous two-currency example, we are now considering increasing the CAD position by US$10,000.

First, we use the marginal-VaR method. We note that β can be obtained from a previous intermediate step. Because we used dollar amounts, this should be adjusted so that β is unitless, that is,

$$\beta = \frac{\Sigma w}{w'\Sigma w} = W \times \frac{\Sigma x}{x'\Sigma x}$$

We have

$$\beta = \$3 \times \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} / (\$0.156^2) = \$3 \begin{bmatrix} 0.205 \\ 0.590 \end{bmatrix} = \begin{bmatrix} 0.615 \\ 1.770 \end{bmatrix}$$

The marginal VaR is now

$$\Delta\text{VaR} = \alpha \frac{\text{cov}(R, R_p)}{\sigma_p} = 1.65 \times \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} / \$0.156 = \begin{bmatrix} 0.0528 \\ 0.1521 \end{bmatrix}$$

As we increase the first position by $10,000, the incremental VaR is

$$(\Delta\text{VaR})' \times a = [0.0528\ 0.1521] \begin{bmatrix} \$10{,}000 \\ 0 \end{bmatrix}$$

$$= 0.0528 \times \$10{,}000 + 0.1521 \times 0 = \$528$$

Next, we compare this with the incremental VaR obtained from a full revaluation of the portfolio risk. Adding $0.01 million to the first position, we find

$$\sigma^2_{p+a} W^2_{p+a} = [\$2.01\ \$1] \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.12^2 \end{bmatrix} \begin{bmatrix} \$2.01 \\ \$1 \end{bmatrix}$$

which gives VaR_{p+a} = $258,267. Relative to the initial VaR_p = $257,738, the exact increment is $529. Note how close the ΔVaR approximation of $528 comes to the true value. The linear approximation is excellent because the change in the position is very small.

Component VaR

In order to manage risk, it would be extremely useful to have a *risk decomposition* of the current portfolio. This is not straightforward because the portfolio volatility is a highly nonlinear function of its components. Taking all individual VaRs, adding them up, and computing their percentage, for instance, is not useful because it completely ignores diversification effects. Instead, what we need is an additive decomposition of VaR that recognizes the power of diversification.

This is why we turn to marginal VaR as a tool to help us measure the contribution of each asset to the existing portfolio risk. Multiply the marginal VaR by the current dollar position in asset or risk factor i, that is,

$$\text{Component VaR}_i = (\Delta\text{VaR}_i) \times w_i W$$

$$= \frac{\text{VaR}\beta_i}{W} \times w_i W = \text{VaR}\beta_i w_i \qquad (5.27)$$

Thus the component VaR indicates how the portfolio VaR would change approximately if the component was deleted from the portfolio. We should note, however, that the quality of this linear approximation improves when the VaR components are small. Hence this decomposition is more useful with large portfolios, which tend to have many small positions.

We now show that these component VaRs precisely add up to the total portfolio VaR. The sum is

$$\text{C VaR}_1 + \text{CVaR}_2 + \ldots + \text{CVaR}_N = \text{VaR}\left(\sum_{i=1}^{N} w_i\beta_i\right) = \text{VaR} \qquad (5.28)$$

because the term between parentheses is simply the beta of the portfolio with itself, which is unity.[2] Thus we established that these *component* VaR measures add up to the total VaR. We have an additive measure of portfolio risk that reflects correlations. Components with a negative sign act as a hedge against the remainder of the portfolio. In contrast, components with a positive sign increase the risk of the portfolio.

> **Component VaR** A partition of the portfolio VaR that indicates how much the portfolio VaR would change approximately if the given component was deleted. By construction, component VaRs sum to the portfolio VaR.

[2] This can be proved by expanding the portfolio variance into $\sigma_p^2 = w_1 \text{cov}(R_1, R_p) + w_2 \text{cov}(R_2, R_p) + \cdots = w_1(\beta_1\sigma_p^2) + w_2(\beta_2\sigma_p^2) + \cdots = \sigma_p^2(\sum_{i=1}^{N} w_i\beta i)$. Therefore, the term between parentheses must be equal to 1.

The component VaR can be simplified further. Taking into account the fact that β_i is equal to the correlation ρ_i times σ_i divided by the portfolio σ_p, we can write

$$\text{CVaR}_i = \text{VaR}w_i\beta_i = (\alpha\sigma_p W)w_i\beta_i = (\alpha\sigma_i w_i W)\rho_i = \text{VaR}_i\rho_i \quad (5.29)$$

This conveniently transforms the individual VaR into its contribution to the total portfolio simply by multiplying it by the correlation coefficient.

Finally, we can normalize by the total portfolio VaR and report

Percent contribution to VaR of component

$$i = \frac{\text{CVaR}_i}{\text{VaR}} = w_i\beta_i \quad (5.30)$$

VaR systems can provide a breakdown of the contribution to risk using any desired criterion. For large portfolios, component VaR may be shown by type of currency, by type of asset class, by geographic location, or by business unit. Such detail is invaluable for drill-down exercises, which enable users to control their VaR.

Example 5.1 (continued)

Continuing with the previous two-currency example, we find the component VaR for the portfolio using $\text{CVaR}_i = \Delta\text{VaR}_i\, x_i$, that is,

$$\begin{bmatrix} \text{CVaR}_1 \\ \text{CVaR}_2 \end{bmatrix} = \begin{bmatrix} 0.0528 \times \$2 \text{ million} \\ 0.1521 \times \$1 \text{ million} \end{bmatrix} = \begin{bmatrix} \$105{,}630 \\ \$152{,}108 \end{bmatrix}$$

$$= \text{VaR} \times \begin{bmatrix} 41.0\% \\ 59.0\% \end{bmatrix}$$

We verify that these two components indeed sum to the total VaR of \$257,738. The largest component is due to the EUR, which has the highest volatility. Both numbers are positive, indicating that neither position serves as a net hedge for the portfolio. Note that the percentage contribution to VaR also could have been obtained as

$$\begin{bmatrix} \text{CVaR}_1/\text{VaR} \\ \text{CVaR}_2/\text{VaR} \end{bmatrix} = \begin{bmatrix} w_1\beta_1 \\ w_2\beta_2 \end{bmatrix} = \begin{bmatrix} 0.667 \times 0.615 \\ 0.333 \times 1.770 \end{bmatrix}$$

$$= \begin{bmatrix} 41.0\% \\ 59.0\% \end{bmatrix}$$

Next, we can compute the change in the VaR if the euro position is set to zero and compare with the preceding result. Since the portfolio has only two assets, the new VaR without the EUR position is simply the VaR of the CAD component, $\text{VaR}_1 = \$165{,}000$. The incremental VaR of the EUR position is $(\$257{,}738 - \$165{,}000) = \$92{,}738$. The component VaR of \$152,108 is higher, although of the same order of magnitude. The approximation is not as good as before because there are only two assets in the portfolio, which individually account for a large proportion of the total VaR. We would expect a better approximation if the VaR components are small relative to the total VaR.

5.3 SUMMARY

Figure 5.4 presents a graphic summary of VaR tools for our two-currency portfolio. The graph plots the portfolio VaR as a function of the amount invested in this asset, the euro. At the current position of \$1 million, the portfolio VaR is \$257,738.

The marginal VaR is the change in VaR owing to an addition of \$1 in EUR, or 0.0528; this represents the slope of the straight line that is tangent to the VaR curve at the current value.

The incremental VaR is the change in VaR owing to the deletion of the euro position, which is \$92,738 and is measured along the curve. This is approximated by the component VaR, which is simply the marginal VaR times the current position of \$1 million, or \$152,108. The latter is measured along the straight line that is tangent to the VaR curve. The graph illustrates that the component VaR is only an approximation of the incremental VaR. These component VaR measures add up to the total portfolio VaR, which gives a quick decomposition of the total risk.

The graph also shows that the best hedge is a net zero position in the euro. Indeed, the VaR function attains a minimum when the position in the euro is zero.

The results are summarized in Table 5.1. This report gives not only the portfolio VaR but also a wealth of information for risk

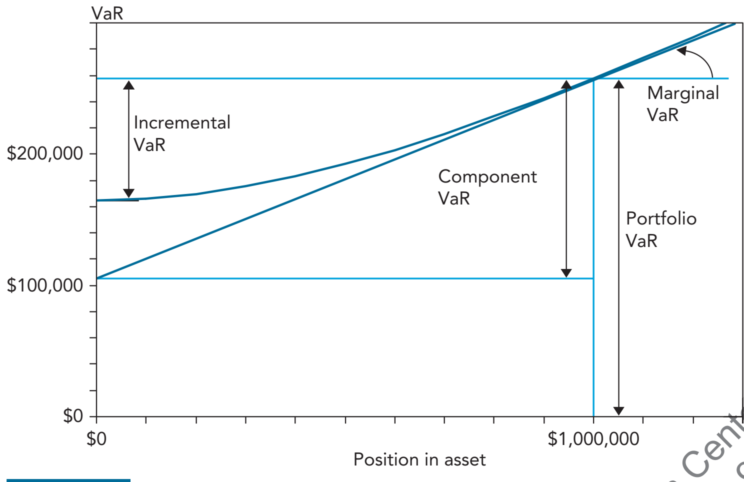


Figure 5.4 VaR decomposition.

Table 5.1 VaR Decomposition for Sample Portfolio

Currency	Current Position, x_i or w_iW	Individual VaR, $\text{VaR}_i = \alpha\sigma_i w_i W$	Marginal VaR, $\Delta\text{VaR}_i = \text{VaR}\beta_i/W$	Component VaR, $\text{CVaR}_i = \Delta\text{VaR}_i x_i$	Percent Contribution, CVaR_i/VaR
CAD	$2 million	$165,000	0.0528	$105,630	41.0%
EUR	$1 million	$198,000	0.1521	$152,108	59.0%
Total	$3 million				
Undiversified **VaR**		$363,000			
Diversified **VaR**				$257,738	100.0%

managers. For instance, the marginal VaR column can be used to determine how to reduce risk. Since the marginal VaR for the EUR is three times as large as that for the CAD, cutting the position in the EUR will be much more effective than cutting the CAD position by the same amount.

5.4 EXAMPLES

This section provides a number of applications of VaR measures. The first example illustrates a risk report for a global equity portfolio. The second shows how VaR could have been used to dissect the Barings portfolio.

A Global Portfolio Equity Report

To further illustrate the use of our VaR tools, Table 5.2 displays a risk management report for a global equity portfolio. Here, risk is measured in relative terms, that is, relative to the benchmark portfolio. The current portfolio has an annualized tracking error volatility σ_p, of 1.82 percent per annum. This number can be translated easily into a VaR number using $\text{VaR} = \alpha\sigma_p W$. Hence we can deal with VaR or more directly with σ_p.

Positions are reported as deviations in percent from the benchmark in the second column. Since the weights of the benchmark and of the current portfolio must sum to one, the deviations must sum to zero. Traditional portfolio reporting systems only provide information about current positions for the portfolio. The position data, however, could be used to provide detailed information about risk.

The next columns report the individual risk, marginal risk, and percentage contribution to total risk. Positions contributing to more than 5 percent of the total are called *Hot Spots*.[3] The table shows that two countries, Japan and Brazil, account for more than 50 percent of the risk. This is an important but not intuitive result because the positions in these markets, displayed in the first column, are not the largest in terms of weights.

In fact, the United States and United Kingdom, which have the largest deviations from the index, contribute to only 20 percent of the risk. The contributions of Japan and Brazil are high because of their high volatility and correlations with the portfolio.

To control risk, we turn to the "Best Hedge" column. The table shows that the 4.5 percent overweight position in Japan should be decreased to lower risk. The optimal change is a decrease of 4.93 percent, after which the new volatility will have decreased from the original value of 1.82 to 1.48 percent. In contrast, the 4.0 percent overweight position in Canada has little impact on the portfolio risk.

This type of report is invaluable to control risk. In the end, of course, portfolio managers add value by judicious bets on markets, currencies, or securities. Such VaR tools are useful, however, because analysts now can balance their return forecasts against risk explicitly.

Barings: An Example in Risks

Barings' collapse provides an interesting application of the VaR methodology. Leeson was reported to be long about $7.7 billion worth of Japanese stock index (Nikkei) futures and short $16 billion worth of Japanese government bond (JGB) futures. Unfortunately, official reports to Barings showed "nil" risk because the positions were fraudulent.

If a proper VaR system had been in place, the parent company could have answered the following questions: What was Leeson's actual VaR? Which component contributed most to VaR? Were the positions hedging each other or adding to the risk?

[3] Hot Spots is a trademark of Goldman Sachs.

Table 5.2 Global Equity Portfolio Report

Country	Current Position (%) w_i	Individual Risk $w_i\sigma_i$	Marginal Risk β_i	Percent Contribution to Risk, $w_i\beta_i$	Best Hedge (%)	Volatility at Best Hedge
Japan	4.5	0.96%	0.068	**31.2**	−4.93	1.48%
Brazil	2.0	1.02%	0.118	**22.9**	−1.50	1.66%
U.S.	−7.0	0.89%	−0.019	**13.6**	3.80	1.75%
Thailand	2.0	0.55%	0.052	**10.2**	−2.30	1.71%
U.K.	−6.0	0.46%	0.035	**7.0**	2.10	1.80%
Italy	2.0	0.79%	−0.011	**6.8**	−2.18	1.75%
Germany	2.0	0.35%	0.019	3.7	−2.06	1.79%
France	−3.5	0.57%	−0.009	3.4	1.18	1.81%
Switzerland	2.5	0.39%	0.011	2.6	−1.45	1.81%
Canada	4.0	0.49%	0.001	1.5	−0.11	1.82%
South Africa	−1.0	0.20%	0.008	−0.7	−0.65	1.82%
Australia	−1.5	0.24%	0.014	−2.0	−1.89	1.80%
Total	0.0			100.0		
Undiversified risk		6.91%				
Diversified risk	1.82%					

Source: Adapted from Litterman (1996).

Table 5.3 Barings' Risks

	Risk % σ	Correlation Matrix R		Covariance Matrix Σ		Positions ($ millions) x	Individual VaR $\alpha\sigma x$
10-year JGB	1.18	1	−0.114	0.000139	−0.000078	($16,000)	$310.88
Nikkei	5.83	−0.114	1	−0.000078	0.003397	$7,700	$740,51
Total						$8,300	$1051.39

		Total VaR Computation	Marginal VaR β_i for $1 million			
Asset i	$(\Sigma x)_i$	$x_i'(\Sigma x)_i$	$(\Sigma x)_i/\sigma_p^2$	β_i VaR	Component VaR $\beta_i x_i$VaR	Percent Contribution
10-yr JGB	−2.82	45138.8	−0.0000110	($0.00920)	$147.15	17.6%
Nikkei	27.41	211055.1	0.0001070	$0.08935	$688.01	82.4%
Total		256193.8			$835.16	100.0%
Risk $= \sigma_p$		506.16				
VaR $= \alpha\sigma_p$		$835.16				

The top panel of Table 5.3 displays monthly volatility measures and correlations for positions in the 10-year zero JGB and the Nikkei Index. The correlation between Japanese stocks and bonds is negative, indicating that increases in stock prices are associated with decreases in bond prices or increases in interest rates. The next column displays positions that are reported in millions of dollar equivalents.

To compute the VaR, we first construct the covariance matrix Σ from the correlations. Next, we compute the vector Σx, which is in the first column of the bottom panel. For instance, the -2.82 entry is found from $\sigma_1^2 x_1 + \sigma_{12} x_2 = 0.000139 \times (-\$16{,}000) + (-0.000078) \times \$7700 = -2.82$. The next column reports $x_1(\Sigma x)_1$ and $x_2(\Sigma x)_2$, which sum to the total portfolio variance of 256,193.8, for a portfolio volatility of $\sqrt{256{,}194} = \$506$ million. At the 95 percent confidence level, Barings' VaR was $1.65 \times \$506$, or \$835 million.

This represents the worst monthly loss at the 95 percent confidence level under normal market conditions. In fact, Leeson's total loss was reported at \$1.3 billion, which is comparable to the VaR reported here. The difference is because the position was changed over the course of the 2 months, there were other positions (such as short options), and also bad luck. In particular, on January 23, 1995, one week after the Kobe earthquake, the Nikkei Index lost 6.4 percent. Based on a monthly volatility of 5.83 percent, the daily VaR of Japanese stocks at the 95 percent confidence level should be 2.5 percent. Therefore, this was a very unusual move—even though we expect to exceed VaR in 5 percent of situations.

The marginal risk of each leg is also revealing. With a negative correlation between bonds and stocks, a hedged position typically would be long the two assets. Instead, Leeson was short the bond market, which market observers were at a loss to explain. A trader said, "This does not work as a hedge. It would have to be the other way round."[4] Thus Leeson was increasing his risk from the two legs of the position.

This is formalized in the table, which displays the marginal VaR computation. The β column is obtained by dividing each element of Σx by $x'\Sigma x$, for instance, -2.82 by 256,194 to obtain -0.000011. Multiplying by the VaR, we obtain the marginal change in VaR from increasing the bond position by \$1 million, which is $-\$0.00920$ million. Similarly, increasing the stock position by \$1 million increased the VaR by \$0.08935.

Overall, the component VaR owing to the total bond position is \$147.15 million; that owing to the stock position is \$688.01 million. By construction, these two numbers add up to the total VaR of \$835.16 million. This analysis shows that most of the risk was due to the Nikkei exposure and that the bond position, instead of hedging, made things even worse. As Box 5.1 shows, however, Leeson was able to hide his positions from the bank's VaR system.

[4] *Financial Times*, March 1, 1995.

BOX 5.1 BARINGS' RISK MANAGEMENT

The Barings case is a case in point of lack of trader controls. A good risk management system might have raised the alarm early and possibly avoided most of the \$1.3 billion loss.

Barings had installed in London a credit-risk management system in the 1980s. The bank was installing a market-risk management system in its London offices. The system, developed by California-based Infinity Financial Technology, has the capability to price derivatives and to support VaR reports. Barings' technology, however, was far more advanced in London than in its foreign branches. Big systems are expensive to install and support for small operations, which is why the bank relied heavily on local management.

The damning factor in the Barings affair was Leeson's joint responsibility for front- and back-office functions, which allowed him to hide trading losses. In July 1992, he created a special "error" account, numbered 88888, that was hidden from the trade file, price file, and London gross file. Losing trades and unmatched trades were parked in this account. Daily reports to Barings' Asset and Liability Committee showed Leeson's trading positions on the Nikkei 225 as fully matched. Reports to London therefore showed no risk. Had Barings used internal audits to provide independent checks on inputs, the company might have survived.

5.5 VAR TOOLS FOR GENERAL DISTRIBUTIONS

So far we have derived analytical expressions for these VaR tools assuming a normal distribution. These results can be generalized. In Equation (5.1), the portfolio return is a function of the positions on the individual components $R_p = f(w_1, \ldots, w_N)$. Multiplying all positions by a constant k will enlarge the portfolio return by the same amount, that is,

$$kR_p = f(kw_1, \ldots, kw_N) \tag{5.31}$$

Such function is said to be *homogeneous of degree one*, in which case we can apply *Euler's theorem*, which states that

$$R_p = f(w_1, \ldots, w_N) = \sum_{i=1}^{N} \frac{\partial f}{\partial w_i} w_i \tag{5.32}$$

The portfolio VaR is simply a realization of a large dollar loss. Setting R_p to the portfolio VaR gives:

$$\begin{aligned} \mathrm{VaR} &= \sum_{i=1}^{N} \frac{\partial \mathrm{VaR}}{\partial w_i} \times w_i = \sum_{i=1}^{N} \frac{\partial \mathrm{VaR}}{\partial x_i} \times x_i \\ &= \sum_{i=1}^{N} (\Delta \mathrm{VaR}_i) \times x_i \end{aligned} \tag{5.33}$$

Table 5.4 Risk-Minimizing Position

Asset	Original Position, w_i	Marginal VaR, ΔVaR_i	Final Position, w_i	Marginal VaR, ΔVaR_i	Beta β_i
CAD	66.67%	0.0528	85.21%	0.0762	1.000
EUR	33.33%	0.1521	14.79%	0.0762	1.000
Total	100.00%		100.00%		
Diversified **VaR**	$257,738		$228,462		
Standard deviation	15.62%		13.85%		

This shows that the decomposition in Equation (5.28) is totally general. With a normal distribution, the marginal VaR is $\Delta\text{VaR}_i = \beta_i(\alpha\sigma_p)$, which is proportional to β_i. This analytical result also holds for *elliptical distributions*. In these cases, marginal VaR can be estimated using the sample beta coefficient, which uses all the sample information, such as the portfolio standard deviation, and as a result should be precisely measured.

Consider now another situation where the risk manager has generated a distribution of returns $R_{p,1}, \ldots, R_{p,T}$, and cannot to approximate it by an elliptical distribution perhaps because of an irregular shape owing to option positions. VaR is estimated from the observation R_p^*. One can show that applying Euler's theorem gives

$$R_p^* = \sum_{i=1}^{N} E(R_i \mid R_p = R_p^*)w_i \tag{5.34}$$

where the $E(\cdot)$ term is the expectation of the risk factor conditional on the portfolio having a return equal to VaR.[5] Thus CVaR_i could be estimated from the decomposition of R^* into the realized value of each component.

Such estimates, however, are less reliable because they are based on one data point only. Another solution is to examine a window of observations around R^* and to average the realized values of each component over this window.

5.6 FROM VaR TO PORTFOLIO MANAGEMENT

From Risk Measurement to Risk Management

Marginal VaR and component VaR are useful tools, best suited to small changes in the portfolio. This can help the portfolio manager to decrease the risk of the portfolio. Positions should be cut first where the marginal VaR is the greatest, keeping portfolio constraints satisfied. For example, if the portfolio needs to be fully invested, some other position, with the lowest marginal VaR, should be added to make up for the first change.

This process can be repeated up to the point where the portfolio risk has reached a global minimum. At this point, all the marginal VaRs, or the portfolio betas, must be equal:

$$\Delta\text{VaR}_i = \frac{\text{VaR}}{W} \times \beta_i = \text{constant} \tag{5.35}$$

Table 5.4 illustrates this process with the previous two-currency portfolio. The original position of $2 million in CAD and $1 million in EUR created a VaR of $257,738, or portfolio volatility of 15.62 percent. The marginal VaR is 0.1521 for the EUR, which is higher than for the CAD.

As a result, the EUR position should be cut first while adding to the CAD position. The table shows the final risk-minimizing position. The weight on the EUR has decreased from 33.33 to 14.79 percent. The portfolio volatility has been lowered from 15.62 to 13.85 percent, which is a substantial drop. We also verify that the betas of all positions are equal when risk is minimized.

From Risk Management to Portfolio Management

The next step is to consider the portfolio expected return as well as its risk. Indeed, the role of the portfolio manager is to choose a portfolio that represents the best combination of expected return and risk. Thus we are moving from *risk management* to *portfolio management*. We will consider each portfolio in a graph that plots its expected return against its risk, as shown in Figure 5.5.

Define E_p as the expected return on the portfolio. This is a linear combination of the expected returns on the component positions, that is,

$$E_p = \sum_{i=1}^{N} w_i E_i \tag{5.36}$$

[5] For proofs, see Tasche (2000) or Hallerbach (2003).

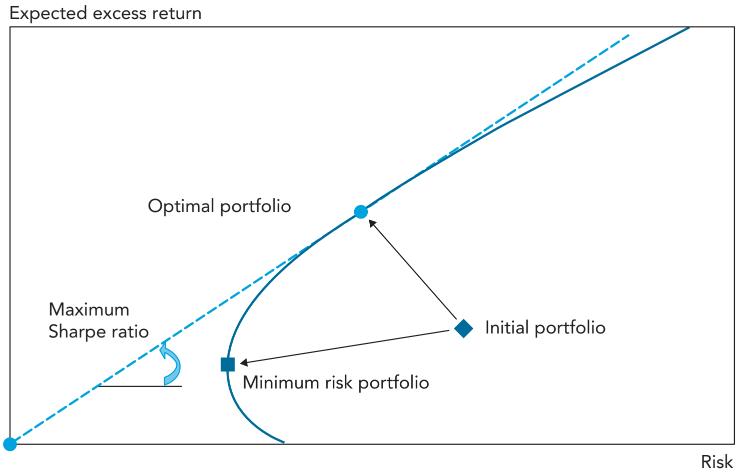


Figure 5.5 **From VaR to portfolio management.**

For simplicity, all returns are defined in excess of the risk-free rate. In the figure, this translates all the points down by the same amount so that the risk-free asset is at the origin.

We then can define the best portfolio combinations as the portfolios that minimize risk for varying levels of expected return. This defines the *efficient frontier*, which is shown as a solid line in Figure 5.5.

Suppose now that the objective function is to maximize the ratio of expected return to risk. This *Sharpe ratio* is

$$SR_p = \frac{E_p}{\sigma_p} \tag{5.37}$$

More generally, this could be written with VaR in the denominator.

How do we move from the current position to this optimal portfolio? The preceding section showed how to move the portfolio from its original position to the *global minimum-risk* portfolio. This portfolio, however, does not take expected returns into account.

We now wish to increase the portfolio expected return as well, moving to the portfolio with the highest Sharpe ratio. This portfolio is on the efficient set and maximizes the slope of the tangent from the origin. We call this portfolio the *optimal portfolio*. At this point, the ratio of all expected returns to marginal VaRs must be equal. This also can be written in terms of the excess expected return for each asset divided by its beta relative to the optimized portfolio. At the optimum,

$$\frac{E_i}{\Delta VaR_i} = \frac{E_i}{\beta_i} = \text{constant} \tag{5.38}$$

Note that this is simply a restatement of the *capital asset pricing model*, which states that the market portfolio must be mean-variance efficient. Roll (1977) showed that the efficiency of any portfolio implies that the expected return on any component asset must be proportional to its beta relative to this portfolio, that is,

$$E_i = E_m\beta_i \tag{5.39}$$

Thus, for each asset, the ratio between the excess return E_i and the beta must be constant.

Table 5.5 shows our two-currency portfolio, for which we assumed that $E_1 = 8$ percent and $E_2 = 5$ percent. The original position has a Sharpe ratio of 0.448. The ratio of E_i/β_i is 0.1301 for CAD, which is greater than the 0.0282 value for EUR. This implies that the CAD position should be increased to improve portfolio performance. Indeed, at the optimum, the CAD weight has increased from 66.67 to 90.21 percent. The portfolio Sharpe ratio has increased substantially from 0.448 to 0.551. We verify

Table 5.5 **Risk and Return–Optimizing Position**

Asset	Expected Return E_i	Original Position w_i	Beta β_i	Ratio E_i/β_i	Final Position w_i	Beta β_i	Ratio E_i/β_i
CAD	8.00%	66.67%	0.615	0.1301	90.21%	1.038	0.0771
EUR	5.00%	33.33%	1.770	0.0282	9.79%	0.649	0.0771
Total		100.00%			100.00%		
Diversified **VaR**		$257,738			$230,720		
Standard deviation		15.62%			13.98%		
Expected return		7.00%			7.71%		
Sharpe ratio		0.448			0.551		

that the ratios E_i/β_i are identical for the two assets at the optimum. The same values of 0.0771 indicate that there is no reason to deviate from the final allocation.

CONCLUSIONS

This chapter has shown how to measure and manage risk using analytical methods based on the standard deviation. Such methods apply when risk factors have distributions that are jointly normal or, more generally, elliptical.

Analytical methods are particularly convenient because they lead to closed-form solutions that are easy to interpret. This is akin to the Black-Scholes model, an analytical model to price options. This model is used widely because it yields powerful insights that can be applied to all options, including those that are computed using numerical methods. Thus the VaR tools developed here for parametric VaR also can be used with non-parametric, simulation-based VaR models.

We have seen that the VaR approach is much richer than the computation of a single risk measure. It provides a framework for managing risk using VaR tools such as marginal VaR and component VaR. These measures can be used to analyze the effect of marginal changes in portfolio composition.

A typical situation is that of a bank trader who has to evaluate whether a proposed trade with a client will increase or decrease the risk of the existing portfolio. Marginal VaR provides useful information to control the risk profile throughout the day. If the trade is risk-decreasing, then the trader should adjust the bid-offer spread to increase the probability that the client will do the trade. On the other hand, a trade that increases risk should be discouraged.

At the end, however, risk is only one component of the portfolio management process. Expected returns must be considered as well. The role of the portfolio manager is to balance increasing risk against increasing expected returns.

This is where VaR methods prove their usefulness. Combining expected profits into a portfolio is an intuitive process because expected returns are additive. In contrast, risk is not additive and is a complicated function of the portfolio positions and risk-factor characteristics. This explains why the battery of VaR tools is useful to manage portfolios better.

6

VaR and Risk Budgeting in Investment Management

Learning Objectives

After completing this reading you should be able to:

- Define risk budgeting.
- Describe the impact of horizon, turnover, and leverage on the risk management process in the investment management industry.
- Describe the investment process of large investors such as pension funds.
- Describe the risk management challenges associated with investments in hedge funds.
- Distinguish among the following types of risk: absolute risk, relative risk, policy-mix risk, active-management risk, funding risk, and sponsor risk.
- Explain the use of VaR to check manager compliance and monitor risk.
- Explain how VaR can be used in the development of investment guidelines and for improving the investment process.
- Describe the risk budgeting process and calculate risk budgets across asset classes and active managers.

Excerpt is Chapter 17 of Value at Risk: The New Benchmark for Managing Financial Risk, *Third Edition, by Philippe Jorion.*

> We have a standard deviation for our total plan, but it only tells you what happened in the past. By contrast, VaR looks forward.
>
> —Director of Chrysler's pension fund

By now, value-at-risk (VaR) has spread well beyond the Wall Street trading departments where it originated. The investment management industry is also discovering the benefits of VaR systems.

Many of the reasons that made VaR successful in the banking industry also apply to asset managers. VaR is a forward-looking measure of the risk profile of a fund based on current positions. The more traditional returns-based approach, in contrast, is purely historical; it does not offer timely measurement of risk.

VaR can be used to measure, control, and manage risk. VaR is comprehensive because it accounts for leverage, volatility, and diversification. VaR is a simple measure of risk that can be explained easily to portfolio managers and investors. VaR systems also can be used to set consistent guidelines that improve over traditional guidelines using limits on notionals or sensitivity measures. As a bonus, comprehensive risk management systems provide some protection against rogue traders, thereby helping to avoid embarrassing financial losses.

This chapter shows how VaR can benefit the investment management industry, which includes mutual funds, pension funds, endowment funds, insurance companies, and hedge funds. This new risk management technique has led to the development of *risk budgeting*. Risk budgeting is the process of allocating and managing risk using a top-down approach to different aspects of the investment process. Risk budgeting builds on VaR measures that are applied to asset classes, asset managers, and even securities.

Risk budgeting is fast spreading as a best-practice method to manage risk. Like VaR, the concept is not new. Like VaR, its main advantage is to provide a top-down, comprehensive, and practical method to manage risk. It provides a dynamic comparison of current risk profiles with prespecified risk budgets.

The first section compares risk measures of proprietary bank trading to the investment management industry. The definition of risk, however, depends on the investment objectives. You will see how to use VaR to monitor and control risk; VaR can be used to manage risk. A particular example of this is the risk-budgeting process.

6.1 VaR APPLICATIONS TO INVESTMENT MANAGEMENT

Sell Side versus Buy Side

The investment management industry usually is called the "buy side" of Wall Street, in contrast with banks, the "sell side" that developed VaR. Whereas VaR has been widely, and rather quickly, accepted by the banking industry, it has spread more slowly to the investment management industry. Perhaps this is so because investment management differs in many fundamental respects from the fast-paced trading environment of dealing banks. Table 6.1 compares the characteristics of the buy side with those of the sell side.

Consider first bank trading portfolios, where the horizon is short, turnover rapid, and leverage high. VaR is particularly appropriate for such an environment. In this case, historical measures of risk basically are useless because yesterday's portfolio profile may have nothing to do with today's. In an investment environment, in contrast, the horizon, as measured by the portfolio evaluation period, is much longer, monthly or quarterly. Positions change more slowly.

Bank trading portfolios are also highly leveraged, which makes it particularly important to control their risk. A sequence of

Table 6.1 Risk Management for the Sell and Buy Sides

Characteristic	Sell Side (e.g., Banks)	Buy Side (e.g., Investors)
Horizon	Short-term (1 day, intraday)	Long-term (month, quarter, years)
Turnover	Rapid	Slow
Leverage	High	Low
Risk measures	VaR Stress tests	Asset allocation Tracking error
Risk controls	Position limits VaR limits Stop-loss rules	Diversification Benchmarking Investment guidelines

adverse events easily could bankrupt the institution, as shown by the Barings crisis. In contrast, pension funds, whose positions are guided by a "prudent investor" philosophy, do not allow much leverage. Thus there is a less crucial need to control the downside risk.

In sum, the daily application of VaR measures has become a requirement of bank trading portfolios owing to short horizons, rapid turnover, and high leverage. Risk is controlled through position limits, VaR limits, and stop-loss rules. Although the investment management industry operates with different risk parameters, the proper measurement of risk is also a critical function. This chapter will demonstrate the benefits of VaR methods for the investment management industry.

Investment Process

To understand the requirements of the investment management industry, it is useful to start by describing the investment process of large investors such as pension funds. Generally, this process consists of two steps. In the first step, a consultant provides a strategic, long-term *asset-allocation* study usually based on mean-variance portfolio optimization, that balances off expected return against risk. This study determines the amounts to be invested in various asset classes, for example, domestic stocks, domestic bonds, foreign stocks, foreign bonds, and perhaps additional classes such as emerging markets, real estate, venture capital, and total-return funds, also known as *hedge funds*. The asset allocation relies on *benchmarks*, or *passive indices*, that represent a feasible investment strategy.

In the second step, the fund may delegate the actual management of funds to a stable of *active* managers. These managers are reviewed periodically for performance relative to their benchmark, measured in terms of their *tracking error*. Risk typically is controlled through a list of *investment guidelines* defining the universe of assets they can invest in, with some additional restrictions such as duration, maximum deviations from equity-sector weights, or maximum amounts of foreign currency to hedge or cross-hedge. Generally, risk is measured ex post, that is, from historical data.

Institutions exposed to a diversity of risks, to complex financial instruments, and to changing positions should benefit from VaR risk management systems. Let us see how these criteria apply to the investment management industry.

First, investments are becoming more global in nature, creating a need for risk measures that take diversification into account. Before 1974, for example, few pension funds invested in foreign markets. By now, funds invest all over the world. They also invest in new asset classes, such as hedge funds.

BOX 6.1 LESSONS FROM WISCONSIN

In March 1995, the State of Wisconsin Investment Board, which controls over \$34 billion in assets, revealed that it had lost \$95 million on currency and interest-rate swaps. While the loss was small in relation to the asset pool, it led to great embarrassment.

Of the total loss, \$35 million came from just one contract, an interest-rate swap that paid

$$\$10 \text{ million} \times (2.95 \text{ percent} - \text{MexSpread})/2.95 \text{ percent}$$

where MexSpread was defined as the yield spread between Mexican and U.S. government bonds. Apparently, the staff had not done a proper sensitivity analysis of the value of the swap and thought that the amount at risk was only \$10 million. In fact, it was much greater owing to the leverage effect induced by the denominator. This loss would have been avoided had the swap been marked to market or, even better, evaluated with a VaR method.

Second, financial instruments are becoming more complex over time. This creates a need for stronger, centralized risk management systems.

In practice, however, institutional investors seem to have too many risk systems. Risk measures usually are based on historical tracking error. Risk controls include "prudent investor" rules based on diversification principles, benchmarking, and investment guidelines. Such systems can have serious flaws, as illustrated in Box 6.1. In contrast, VaR provides a simple, transparent, and consistent measure of overall risk.

Third, most investment portfolios are dynamic, with changing positions. Because the assets of the fund typically are dispersed over a number of managers, it is difficult to create a current picture of the overall risk of the fund. During a quarter, for instance, many fund managers may have increased their exposure to one particular industry. Taken separately, these risks may be acceptable, but as a whole, they may amount to an unsuspected large bet on one source of risk. In addition, money managers sometimes change their investment strategy, either deliberately or inadvertently. If so, the fund should be able to detect and correct such changes quickly. This explains why VaR-based, forward-looking risk measurement systems are also essential to the investment management industry.

Hedge Funds

Hedge funds, however, pose special risk measurement problems. This group is very heterogeneous. Most hedge funds have

leverage. Some groups have greater turnover than traditional investment managers. Long Term Capital Management is an extreme example of a hedge fund that went nearly bankrupt owing to its huge leverage. Such hedge funds are more akin to the trading desks of investment banks than to those of pension funds. As such, they should use similar risk management systems.

Another category of funds, however, invests in *illiquid* assets, such as convertible bonds, which are traded infrequently, even within a month. When this is the case, risk measures based on monthly returns give a misleading picture of risk because the closing *net asset value* (NAV) does not reflect recent transaction prices. This creates two types of biases.

First, correlations with other asset classes will be artificially lowered, giving the appearance of low systematic risk. This can be corrected using enlarged regressions with additional lags of the market factors and summing the coefficients across lags.[1]

Second, volatility will be artificially lowered, giving the appearance of low total risk. Such illiquidity, however, will show up in positive serial autocorrelation in returns. Biases in volatility measures can be corrected by taking this autocorrelation into account when extrapolating risk to longer horizons.[2]

Finally, hedge funds can pose special problems owing to their *lack of transparency*. Many hedge funds refuse to reveal information about their positions for fear of others taking advantage of this information. For clients, however, this makes it difficult to measure the risk of their investment both at the hedge-fund level and in the context of their broader portfolio.

6.2 WHAT ARE THE RISKS?

First, we have to define the risks in investment management. Risk can be clearly defined for a bank trader. It is the risk of loss on the marked-to-market position. Investment asset managers, however, can have different perceptions of risk.

Absolute and Relative Risks

Risk can be defined as the possibility of losses measured in the base currency, dollar or other. This is the most common definition of risk. For managers who have a mandate to beat a benchmark, however, risk must be measured in relative terms. We can distinguish between two definitions:

- *Absolute risk*, which is the risk of a dollar loss over the horizon. This is the usual definition of risk in a trading environment. Sometimes this is called *asset risk*. The relevant rate of return is R_{asset}.
- *Relative risk*, which is the risk of a dollar loss in a fund relative to its benchmark. This shortfall is measured as the dollar difference between the fund return and that of a like amount invested in the benchmark. The relevant return is the *tracking error* $E = R_{\text{asset}} - R^b$, which is the excess return of the asset over the benchmark. If this is normally distributed, VaR can be measured from the standard deviation of the tracking error σ_E as VaR $= \alpha W_0 \sigma_E$.

Policy Mix and Active Management Risk

Consider next a fund that allocates its investment to a pool of active managers in various asset classes. The absolute performance of the fund can be broken down into two components, one owing to policy (or benchmark) choice and the other to active management. Hence total asset risk can be attributed to two sources, the risk of the total policy mix and the risk of active manager deviations from the policy mix:

- *Policy-mix risk*, which is the risk of a dollar loss owing to the policy mix selected by the fund. Since the policy mix generally can be implemented by investing in passive funds, this risk represents that of a passive strategy.
- *Active-management risk*, which is the risk of a dollar loss owing to the total deviations from the policy mix. This represents the summation of profits or losses across all managers relative to their benchmark. Thus there may be diversification effects across managers, depending on whether they have similar styles or not. In addition, the current asset-allocation mix may deviate temporarily from the policy mix.

The absolute risk can be measured from fund returns and can be defined as

$$R_{\text{asset}} = \sum_i w_i R_i \tag{6.1}$$

Where w_i is the weight on fund i with return R_i. This return can be decomposed into

$$R_{\text{asset}} = R_{\text{policy mix}} + R_{\text{active mgt.}} = \sum_i w_i^b R_i^b + \sum_i (w_i R_i - w_i^b R_i^b) \tag{6.2}$$

where R_i^b represents the return on the benchmark for fund i, and w_i^b is its policy weight. If the pension plan deviates from its policy mix ($w_i \neq w_i^b$), the active-management portion can be decomposed further into a term that represents policy decisions and manager performance.

[1] This correction was proposed originally by Dimson (1979). Asness et al. (2001) show that many hedge funds have greater systematic risk with this correction.

[2] See Getmansky et al. (2004). Positive autocorrelation increases long-term risk more quickly than does the usual square-root-of-time rule.

The funds' total VaR can be obtained from the policy-mix VaR, the active-management VaR, and a cross-product term. As an example, the Ontario Teachers' Pension Plan Board (OTPPB) estimates that its annual VaR at the 99 percent level of confidence can be decomposed as follows (in percent of the initial fund value):

Source of Risk	VaR
Policy-mix VaR	19.6%
Active-mgt. VaR	1.6%
Asset VaR	19.3%

This table points to a number of interesting observations. First, most of the risk is due to the policy mix. This is a general result owing to Brinson et al. (1986), who demonstrated that most of the variation in portfolio performance can be attributed to the choice of asset classes. In other words, the choice of mix of stocks and bonds will have more effect on the portfolio performance than the choice of a particular equity or bond manager.

The second interesting result is that the active-management VaR is rather small. Apparently, the fund diversifies away much of the risk of managers deviating from their benchmarks through a careful choice of various styles or many managers. Another explanation is that most of the assets are invested in indexed or closely indexed funds.

Finally, the table shows that the policy-mix VaR and active-management VaR do not add up to the total-asset VaR. In fact, there is a slightly negative correlation between the two, leading to a lower overall asset VaR. If this occurs, active managers could take greater deviations from their benchmark without affecting the plan's total VaR.

This analysis is a good example of insights created by a VaR analysis. Such decomposition can help sponsors to make more informed decisions.

Funding Risk

Focusing on the volatility of assets alone, however, may not be appropriate if the assets are supposed to cover fixed liabilities. Notably, a pension fund with *defined benefits* promises a stream of fixed payments to retirees. If the assets are not sufficient to cover these liabilities, the shortfall will have to be made up by the fund's owner. On the other hand, *defined-contribution* plans put the risk on the employees. In other words, risk should be viewed in an *asset/liability management* (ALM) *framework*. We can define *funding risk* as the risk that the value of assets will not be sufficient to cover the liabilities of the fund.

The relevant variable is the *surplus S*, defined as the difference between the value of assets A and liabilities L. The change then is $\Delta S = \Delta A - \Delta L$. Normalizing by the initial value of assets, we have

$$R_S = \frac{\Delta S}{A} = \frac{\Delta A}{A} - \frac{\Delta L}{L}\frac{L}{A} = R_{\text{asset}} - R_{\text{liabilities}}\frac{L}{A} \qquad (6.3)$$

where $R_{\text{liabilities}}$ is the rate of return on liabilities.

While the value of assets can be measured by marking to market, liabilities are more difficult to evaluate. For pension funds, this represents *accumulated-benefit obligations*, which measure the present value of pension benefits owed to employees discounted at an appropriate interest rate. When liabilities consist mainly of nominal payments, their value in general will behave like a short position in a long-term bond. Thus decreases in interest rates, while beneficial for equities on the asset side, can increase even more the value of liabilities, thereby negatively affecting the surplus. If liabilities are indexed to inflation, they behave like inflation-protected bonds.

The minimum-risk position then corresponds to an *immunized* portfolio, where the duration of the assets matches that of the liabilities. In practice, it may not be possible to immunize the liabilities completely if the existing pool of long-term bonds is insufficient. More generally, immunization carries an opportunity cost if other asset classes generate greater returns over time.

This funding risk represents the true long-term risk to the owner of the fund. If the surplus turns negative, it will have to provide additional contributions to the fund. Sometimes this is called *surplus at risk* (SAR). An example is given in Box 6.2.

As an example, consider a hypothetical pension plan. Call it Public Employee Retirement Fund (PERF). PERF has \$1000 in assets and \$900 in liabilities, for a surplus S of \$100 million. The duration of liabilities is 15 years; this high number is typical for pension funds. Assume that the expected return on the surplus, scaled by assets, is 5 percent. This translates into an expected growth of \$50 million over 1 year, creating an expected surplus of \$150 million. For Canadian pension funds, the typical volatility of the surplus is 9.4 percent, leading to an annual VaR of 22 percent, or \$220 million at the 99 percent confidence level.[3] Taking the deviation between the expected surplus and VaR, we find that there is a 1 percent probability that over the next year the surplus will turn into a deficit of \$70 million or more. The trade-off between this number and an expected surplus growth of \$50 million defines the risk profile of the fund. If acceptable, risk budgeting then allocates the SAR of \$220 million to different aspects of the investment process.

[3] This assumes a normal distribution, with $\alpha = 2.33$. Ambachtsheer (2002) provides an interesting analysis of the risk profile of a sample of Canadian and U.S. pension funds. The average surplus volatility is 8.1 and 18.1 percent for these funds, respectively. Over the 1997–2001 period, however, the growth in the surplus has been very small, which reflects poorly performing stock markets.

BOX 6.2 SURPLUS AT RISK AT OTPPB

The Ontario Teachers' Pension Plan Board (OTPPB) has been at the forefront of applying VaR techniques among institutional investors. OTPPB is the biggest pension fund in Canada, with about C$90 billion (US$78 billion) in assets in 2005.

The plan is required to deliver *defined benefits* to Ontario's teachers during their retirement years. Its stated objective is to earn a high rate of return, at least as great as the rate of inflation plus 5 percent per annum, while minimizing the risk of a contribution increase. Until 1990, OTPPB could invest in Ontario bonds only. Starting in 1990, the plan embarked on an ambitious drive to expand into broader asset classes, with the guidance of a risk management system.

The OTPPB has decided on a policy mix of 45 percent equities, 23 percent fixed-income and absolute strategies, and 32 percent inflation-sensitive investments (i.e., commodities, real estate, and real-return bonds). The goal of this mix is to achieve a long-term surplus growth of 1.3 percent per annum. This translates into a surplus VaR of 22 percent, which is also C$20 billion when applied to assets.

In 1996, OTPPB purchased a firmwide risk management system sold by Sailfish that cost about $500,000. Management has access to daily risk reports, and the board receives monthly risk reports. VaR is measured as the worst loss at the 99 percent confidence level over 1 year. Thus we would expect on average a loss worse than C$20 billion in 1 year out of a hundred. Of course, this invites quips about risk managers not likely to be around for a century. The risk managers then patiently explain that these parameters are equivalent to a confidence level of 90 percent over 4 years.* Thus this loss would be expected in one of ten periods of 4 years.

* See De Bever et al. (2000). This transformation assumes normal and i.i.d. returns, in which case the ratio of 90 and 99 percent normal deviates is 1.28/2.33, which multiplied by $\sqrt{4}$ indeed gives a number close to 1.

Sponsor Risk

This notion of surplus risk can be extended to the risk to the owner of the fund, the plan sponsor, who ultimately bears responsibility for the pension fund. One can distinguish between the following risk measures:

- *Cash-flow risk*, which is the risk of year-to-year fluctuations in contributions to the pension fund. Plan sponsors that can absorb greater variations in funding costs, for instance, can adopt a more volatile risk profile.
- *Economic risk*, which is the risk of variation in total economic earnings of the plan sponsor. The surplus risk may be less of a concern, for instance, if falls in the surplus occur in an environment where the firm enjoys greater operating profits.[4]

From the viewpoint of the plan sponsor, risk is measured not only by movements in the assets, or even the surplus, but also by the ultimate effect on the economic value of the firm. Thus pension-plan management should be integrated with the overall financial goals of the plan sponsor. This is in line with the trend toward enterprisewide risk management.

[4] When the pension fund is heavily invested in stocks, the opposite effect will occur. Downturns in the economy will push down the value of the fund's assets, precisely when the company is less able to make contributions. Black (1980) also pointed out that because returns in pension funds are not taxed, corporate bonds should be the preferred investment.

6.3 USING VaR TO MONITOR AND CONTROL RISKS

VaR systems can be used to measure and control market risks. This also applies to the investment management industry. VaR systems allow investors to check that their managers comply with guidelines and to monitor their market risks. *Credit risk* usually is controlled through limits on exposures on a name-by-name basis. VaR systems provide some protection against *operational risk*, which is also controlled by policies and procedures. Such applications are still *passive* or *defensive* in nature.

Using VaR to Check Compliance

The impetus for centralized risk management in the investment management industry came from the realization that the industry is not immune to the "rogue trader" syndrome that has plagued the banking industry. Indeed Box 6.3 explains how the Common Fund lost $138 million from unauthorized trading. This has led to a reorganization of the fund with a centralized risk management function.

The lessons from this loss are applicable to any "manager of managers," that is, a manager who delegates the actual investment decisions to a stable of managers. While rogue traders, fortunately, are rare, minor violations of investment guidelines occur routinely. Some securities may be prohibited because of

BOX 6.3 NONCOMPLIANCE AT THE COMMON FUND

In 1995, the Common Fund, a nonprofit organization that manages about $20 billion on behalf of U.S. schools and universities, announced that it had lost $138 million from unauthorized trading by one of its managers. Apparently, Kent Ahrens, a trader at First Capital Strategies, had deviated from what should have been a safe index-arbitrage strategy between stock index futures and underlying stocks. One day he failed to complete the hedge and lost $250,000. He then tried to trade his way out of this loss but with little success. The growing loss was concealed for 3 years until Ahrens confessed in June 1995.

This loss was all the more disturbing because after the Barings affair, the Common Fund had specifically asked First Capital to demonstrate that a rogue trader could not do the same thing at First Capital. The firm answered that market neutrality was being verified daily. In this case, it seems that proper checks and balances were not in place.

Although the dollar loss was not large compared with the size of the asset pool, it severely damaged the reputation of the Common Fund. Several fund investors left, taking $1 billion with them. The fallout also forced the president of the Common Fund the resign.

In retrospect, the Common Fund realized that running an operation with a large number of fund managers requires strong centralized controls. To prevent such mishaps, the fund created the new position of "independent risk oversight officer." The fund also set up new risk management committees, one of which is at the board level. Its custodian, Mellon Trust, developed online software that checks for violations of investment policies. To reduce operational risk, the fund also cut the number of active managers and custodial agreements.

their risks or for other reasons (e.g., political or religious). Bank custodians, however, indicate that fund managers sometimes trade in and out of unauthorized investments before the client realizes what happened. With monthly reporting, it is hard to catch such movements. Centralized risk management systems, in contrast, can monitor investments in real time.

Such occurrences have moved the pension-fund industry toward centralized risk management. VaR systems provide a central repository for all positions. Independent reconciliation against manager positions makes fraud a lot more difficult. VaR systems also allow users to catch deviations from stated policies quickly.

Using VaR to Monitor Risk

With a VaR system in place, investors can monitor their market risk better. This applies to both passive and active allocations.

Passive allocation, or benchmarking, does not keep risk constant because the composition of the indices can change substantially. The late 1990s, for example, witnessed a high-tech bubble that increased the market capitalization of firms in high-tech industries. As a result, market-capitalization benchmarks such as the Standard & Poor's (S&P) 500 became increasingly exposed to the high-tech industry, which sharply increased the volatility of the indices. Such trends would be picked up by a forward-looking risk measurement system.

Active portfolio management can change the risk profile of the fund. Suppose, for instance, that the investor notices a sudden jump increase in the reported VaR of the fund. The key is to identify the reason for the jump. Several explanations are possible, each requiring different actions.

- *A manager taking more risk*. VaR allows dynamic risk monitoring of managers, who are given a VaR limit or risk budget. Any exceedence of the VaR limit will be flagged and should be examined closely. For instance, if this is an unauthorized trade, the infraction should be corrected at once. Otherwise, the exceedence requires a discussion with the manager. There may be good reasons to increase the risk profile. Perhaps the risk increase is temporary or justified by current conditions. In any event, it is important to understand the reason behind the change.
- *Different managers taking similar bets.* This can happen, for instance, when managers increase their allocation to a particular sector, which is perhaps becoming more attractive or has performed well in the recent past. Because active managers operate in isolation, such a problem can be caught only at the portfolio level. To decrease the portfolio risk, managers can be given appropriate instructions.
- *More volatile markets.* VaR can increase if the current environment becomes more volatile, assuming that time variation in risk is explicitly modeled, such as with GARCH models. The plan sponsor then will have to decide whether it is worth accepting greater volatility. If the risks are deemed to be too large, positions can be cut. Increased volatility, however, often is associated with falls in asset prices leading to correspondingly higher expected returns. Thus the rebalancing decision involves a delicate trade-off between risk and return. As seen in Box 6.4, however, some investors prefer to set up their system so as to smooth out spikes in risk.

More generally, VaR can be reverse engineered to understand where risk is coming from using VaR tools. Measures of marginal

BOX 6.4 SMOOTHING OUT RISK AT OTPPB

The OTPPB risk measurement system is based on historical simulation because of its ability to represent non-normal market movements. The system loads more than 10,000 positions on a daily basis, which are combined with historical data going to January 1987. As of 2005, this represents an expanding window spanning 19 years.

This long history was selected for two reasons. First, the risk managers wanted to include the crash of October 1987 in the sample period so as to model the possibility of a future crash. Second, this long window decreases the weight on recent observations, which smoothes out the volatility process. Shorter windows lead to greater fluctuations in risk measures, which create problems when investment managers are subject to strict risk limits. OTPPB's risk managers indicated that "using a lot of history avoids a potential conflict between risk control and investment strategy."

and component VaR can be used to identify where position changes will have the greatest effect on the total portfolio risk.

This assumes, however, that all the relevant risks are captured by the risk management system. As explained earlier, risk cannot be measured easily for some important asset classes such as real estate, venture capital, and some categories of hedge funds owing to illiquidity. Other series may have very short histories, such as emerging markets, or none at all, such as initial public offerings. In some cases, the missing series can be replaced by a *proxy*, using a mapping approach. The risk manager should be aware of the limitations of the system.

The Role of the Global Custodian

The philosophy behind VaR is centralized risk management. The easiest path to centralization is to use one global custodian only.

This explains why many investors now are aggregating their portfolio holdings with a single custodian. With one global custodian, position reports directly give a consolidated picture of the total exposure of the fund. Custodians become the natural focal point for this analysis because they already maintain position information and have market data. The next level of service is to combine the current position with forward-looking risk measures.

Not all agree, however, that the risk measurement function can be delegated to the custodian. Some larger plans have decided to develop their own internal risk management system. Their rationale is that they have tighter control over risk measures and can better incorporate VaR systems into operations. Larger plans also benefit from economies of scale, spreading the cost of risk management systems over a large asset base, and also require tighter control when their assets are partly managed internally.

These clients are the exception, however. Most investors may be content with risk management reports developed by custodians. Such systems, however, are not cheap to develop. As a result, the trend will be toward fewer custodians that can provide more services. Already, large custodian banks such as Deutsche Bank, JPM Chase, Citibank, and State Street are providing risk management products. State Street, for instance, is already providing a Web-based system, called *VaR Calculator*, that allows users to perform VaR calculations on demand.

The Role of the Money Manager

On the money management side, managers are now under pressure from clients to demonstrate that they have in place a sound risk management system. More and more clients are explicitly asking for risk analysis because they are no longer satisfied with quarterly performance reports only.

Increasingly, clients are asking their managers, "What is your risk management system?" Leading-edge investment managers already have adapted VaR systems into their investment management process. Managers who do not have comprehensive risk management systems put themselves at a serious competitive disadvantage. Indeed, the "Risk Standards" developed in 1996 for institutional investors recommend measuring the risk of the overall portfolio, as well at that of each instrument. The report also notes that manager differentiation increasingly is created by providing risk management services to clients.

6.4 USING VaR TO MANAGE RISKS

VaR systems can be used to manage risk, which is an *active* application. VaR can be used to improve investment guidelines for active managers and to help with the investment process. In theory, VaR also could be used to compute the risk-adjusted performance of investment managers, as is done for bank traders.

Using VaR to Design Guidelines

VaR systems can be used to design better investment guidelines. Managers' guidelines generally are set up in an ad hoc fashion to restrict the universe of assets in which the managers can invest and, to some extent, to control risk. Typically, guidelines include limits on *notionals*, for example, maximum sector weight deviations for equities and maximum currency positions,

or limits on *sensitivities*, such as duration gaps between fixed-income portfolios and their benchmarks.

Banking institutions, however, have learned the hard way that limits on notionals and sensitivities are insufficient. Limits on notionals work best with simple portfolios with no derivatives and leverage. They do not account for variations in risk nor correlations. Limits on sensitivities are an improvement but still have blind spots, such as for hedged portfolios. In contrast, VaR limits are comparable across assets and account for risk, diversification, leverage, and derivatives (provided the system is well designed).

Says Leo de Bever, risk manager at Ontario Teachers, "Typically, you control positions by saying, 'Thou shalt not have more than *X* million of this.' When you do that, you end up with a whole bunch of rules on what you can and cannot do, but not a handle on how much you might lose on any given day."

Another problem is that the spirit of these limits can be skirted with new financial instruments. For example, a manager may not be allowed to trade in futures that may be viewed as too "risky," such as futures contracts. Instead, investments may be allowed in high-grade medium-term notes, often viewed as safe because they have no credit risk. The problem is that these notes can be designed as *structured* notes with as much market risk as futures contracts. Hence detailed guidelines, like government regulations, are one step behind continuously changing financial markets. Traditional guidelines cannot cope well with new instruments or leverage. They also totally ignore correlations.

This is precisely what VaR attempts to measure. Instead of detailed guidelines, plan sponsors could specify that the anticipated volatility of tracking error cannot be more than 3 percent, for instance. Position limits can be set consistently across markets.

Using VaR for the Investment Process

A good risk management system can be used to improve the investment process, starting with the top-level asset-allocation process all the way down to trading decisions for individual stocks.

As explained earlier, the strategic asset-allocation decision is the first and most important step in the investment process for pension funds. It is usually based on a mean-variance optimization that attempts to identify the portfolio with the best risk-return trade-off using a set of long-term forecasts for various asset classes.

In practice, the optimization usually is constrained in an effort to obtain solutions that look "reasonable." This adjustment, however, partly defeats the purpose of portfolio optimization and fails to recognize the effects of marginal adjustments from the selected portfolio.

Since VaR is, after all, perfectly consistent with a mean-variance framework, VaR tools can be used to allocate funds across asset classes. Box 6.5 shows how incremental VaR can yield useful insights into the risk drivers of a fund.

Risk management systems are also useful at the trading level. Portfolio managers are paid to take bets. Presumably, they've developed skills in one dimension of the risk-return space. They are expected to identify expected returns on various investments. While expected returns can be estimated on an individual basis, assessing the contribution of a particular stock to the total portfolio risk is much less intuitive. Even if analysts

BOX 6.5 VAR AND CURRENCY HEDGING

Bankers Trust (now Deutsche Bank) recently provided its RAROC 2020 risk management system to the Chrysler pension fund. The system provides measures of total and incremental VaR for the various asset classes in which the fund is invested. It can be used, among other things, to evaluate the effectiveness of hedging strategies. In particular, the fund was considering adding a currency hedge to protect the currency position of its foreign stock and bond investments.

The RAROC system showed that the "individual" risk of a $250 million currency position was $44 million at the 99 percent level over 1 year, which appears substantial. However, the pension fund realized that the "incremental" contribution to total risk of a passive currency hedge program was only $3 million.* This means that currency risk already was largely diversified in the existing portfolio. The fund decided not to hedge its currency exposure, thereby saving hefty management fees. These savings more than offset the modest cost of the RAROC system, which was priced at around $50,000 per year.

This result parallels the discussion in the academic literature, where currency hedging initially was advocated as a "free lunch," that is, lower risk at no cost.† Indeed, currency hedging reduces the volatility of individual asset returns, but this is not the relevant issue. Absent currency views, what matters is total portfolio risk. Empirically, total risk generally is not much affected by currency hedging if the proportion of assets invested abroad is small. Thus there is not much benefit from currency hedging.

* Incremental risk is the change in risk when the position is dropped.

† As in Pérold and Schulman (1988). Jorion (1989), however, argues that the benefit of hedging must be viewed in the context of total portfolio risk.

could measure the individual risk of the particular stock they are considering, they cannot possibly be aware of the relationships between all existing positions of the fund. This is where VaR systems help.

For each asset to be added to the portfolio, analysts should be given a measure of its marginal VaR. If two assets have similar projected returns, the analyst should pick the one with the lowest marginal VaR, which will lead to the lowest portfolio risk. Assume, for instance, that the analyst estimates that two stocks, a utility and an Internet stock, will generate an expected return of 20 percent over the next year. If the current portfolio is already heavily invested in high-tech stocks, the two stocks will have a very different marginal contribution to the portfolio risk. Say that the utility stock has a portfolio beta of 0.5 against 2.0 for the other stock, leading to a lower marginal VaR for the first stock. With equal return forecasts, the utility stock is clearly the preferred choice. Such analysis is only feasible within the context of a portfoliowide VaR system.

6.5 RISK BUDGETING

Advances in VaR have led to *risk budgeting*, which is spreading rapidly in investment management. This concept is equivalent to a top-down allocation of economic risk capital starting from the asset classes down to the choice of the active manager and even to the level of individual securities.

Budgeting across Asset Classes

Consider again our pension fund, PERF, that needs to allocate \$1000 million to four asset classes, U.S. and foreign stocks and bonds. Table 6.2 displays the average returns, volatilities, and correlations estimated from 1978–2005 data. These are taken as inputs into the portfolio-allocation process.

Assume now that PERF's board of trustees has decided on a total volatility profile for the fund of 10 percent. This translates into an annual VaR of 23.3 percent, or \$233 million, at the 99 percent confidence level. In what follows, we assume normal distributions and compute VaR as $\alpha\sigma w = 2.33 \times 0.10 \times \$1{,}000 = \$233$ million.

Given this risk appetite, the optimal asset allocation is listed in Table 6.3. These optimal weights can be converted into risk budgets, or individual VaRs. For instance, for U.S. stocks, this is $2.33 \times 0.151 \times \$525 = \$184$ million. Note that the risk budgets sum to \$308 million, which is the undiversified VaR. The actual fund VaR is lower, at \$233 million, owing to diversification effects.

The same process can be continued at the next level. Assume that PERF allocates the \$525 million principal in U.S. stocks to two active managers. The two managers are equally good and receive the same amount, \$262.5 million each. They run portfolios with volatility of 16 percent and correlation of 0.78 with each other. This gives a risk budget of $\alpha\sigma W = 2.33 \times 0.16 \times \$262.5 = \$98$ million each. Again, the

Table 6.2 Asset Classes: Risk and Expected Returns

Asset		Expected Return	Volatility	Correlations			
				1	2	3	4
U.S. stocks	1	13.4%	15.1%	1.00			
Non-U.S. stocks	2	12.8%	16.7%	0.54	1.00		
U.S. bonds	3	8.0%	7.3%	0.19	0.11	1.00	
Non-U.S. bonds	4	9.0%	10.9%	0.05	0.52	0.40	1.00

Table 6.3 Risk Budgeting Across Asset Classes (Millions)

Asset	Weight	Volatility	Principal	Risk Budget
U.S. stocks	52.5%	15.1%	\$525	\$184
Non-U.S. stocks	10.4%	16.7%	\$104	\$41
U.S. bonds	12.2%	7.3%	\$122	\$21
Non-U.S. bonds	24.9%	10.9%	\$249	\$63
Portfolio	100.0%	10.0%	\$1000	\$233

Table 6.4 Risk Budgeting Across Active Managers (Millions)

	Inputs		Outputs			
	TEV ω_i	Information Ratio IR_i	Weight x_i	Allocated Principal x_iW	Excess Return $x_i\mu_i$	Relative Risk Budget
Manager 1	6.0%	0.60	55%	$291	2.0%	$40.6
Manager 2	6.0%	0.40	37%	$194	0.9%	$27.1
Index	0.0%	0.00	8%	$40	0.0%	$0.0
Portfolio	4.0%	0.72	100%	$525	2.9%	$48.9

risk budgets sum to an amount that is greater than the risk budget for this asset class owing to diversification effects. The total risk budget is $\sqrt{\$98^2 + \$98^2 + 2 \times 0.78 \times \$98 \times \$98} = \$184$ million.

Each manager then is charged to earn the highest return on these risk units. Another advantage of this approach is that it avoids micromanaging the investment process. As long as managers stay within their risk guidelines, they can execute new transactions without requiring approval of senior management.

Budgeting across Active Managers

This approach can be refined further if we are willing to make assumptions about the expected performance of active managers. For better or for worse, active managers usually are evaluated in terms of their *tracking error* (TE), defined as the active return minus that of the benchmark.[5] Define μ as the expected TE and ω as its volatility (TEV). The *information ratio* then is defined as

$$IR = \mu/\omega \tag{6.4}$$

Managers are commonly evaluated on the basis of their IR. Grinold and Kahn (1995), for example, assert that an IR of 0.50 is "good," meaning in the top quartile of the active managers. Logically, a greater risk budget should be allocated to managers with better performance, as measured by the IR criterion.

The optimization problem for active manager allocation attempts to maximize the IR for the total portfolio subject to a TEV constraint. Define x_i as the fraction invested in manager i, who has a tracking error of ω_i and excess return of μ_i. The value added for the total portfolio p is

$$\mu_p = \sum_i x_i\mu_i = \sum_i x_i\,(IR_i \times \omega_i) \tag{6.5}$$

Now assume that the deviations for each manager are independent of each other.[6] The portfolio TEV is fixed at

$$\omega_p = \sqrt{\sum_i x_i^2\omega_i^2} \tag{6.6}$$

Maximizing the portfolio information ratio subject to a fixed TEV gives the following solution:

$$x_i\omega_i = IR_i\left(\frac{1}{IR_p}\omega_p\right) \tag{6.7}$$

Thus the relative risk budgets should be proportional to the information ratios.

Table 6.4 gives an example. PERF wants to allocate $525 million to a pool of active managers so as to maximize the information ratio of the fund subject to an overall TEV of 4 percent. This is equivalent to a risk budget of $48.9 million. Each manager has a TEV of 6 percent. To achieve an exact TEV of 4 percent, we also need some residual investment in the benchmark, which has a TEV of zero. The fund managers have different capabilities; their IRs are 0.60 and 0.40, respectively. Equation (6.7) gives a solution of 55 percent weight for manager 1, 37 percent for manager 2, and the residual of 8 percent in the index.

This portfolio is expected to return 2.9 percent. With 4 percent TEV, this translates into an information ratio of 2.9/4.0 = 0.72. This is higher than the IR of both managers and is due to the fact that we assumed that the active returns were independent of each other, leading to substantial diversification benefits.[7]

[5] Note that this approach is a second-best because it fails to control total risk. As shown in Roll (1992) and Jorion (2003), active managers who only pay attention to relative risk often end up increasing total risk.

[6] This is a major simplification for what follows. The analysis, however, can be extended to the more realistic case of nonzero correlations using a correlation matrix.

[7] If the IR were the same for each manager, the total IR would be $IR_i\sqrt{2} = 0.707$ or, more generally, for N managers, $IR_i\sqrt{N}$. This is also known as the *law of active management*. In theory, the fund's information ratio would tend to infinity as N grows large. In practice, however, it is difficult to find a very large number of totally unrelated trading strategies.

Such structured, top-down risk allocation adds tremendous rigor to the investment management process. It has been of great value to early adopters, such as the OTPPB. The VaR system has reduced the number of investment rules and has facilitated the closer supervision of risks. OTPPB's focus on risk and diversification has led to greater investment in alternative asset classes. Since 1990, the plan has beaten its benchmark by an average of 4 percent per annum and has provided $16 billion in value added.

CONCLUSIONS

Centralized risk management systems, by now widely adopted on Wall Street, are also taking hold in the investment management industry. Even though institutional investors have a longer-term horizon than bank trading departments, they also greatly benefit from the discipline provided by VaR systems.

Traditionally, risk has been measured using historical returns or as the occurrence of a big loss. While useful for some purposes, these risk measures have severe shortcomings because they are backward-looking. In contrast, VaR provides forward-looking measures of risk, using a combination of current positions with risk forecasts.

When implemented at the level of the total plan, VaR allows improved control of portfolio risk and of managers. It cuts through the maze of diversification rules, benchmark portfolios, and investment guidelines. VaR systems allow analysts to make better risk-return trade-offs. VaR, of course, will not tell you where to invest. The goal is not to eliminate risk but rather to get the just reward for risk that managers elect to take.

Such risk management systems are spreading quickly among institutional investors, changing the face of the industry. They are affecting the custody business, forcing custodians to offer risk management reporting capabilities. Managers are affected, too. Those who do not have a risk management system put themselves at a serious competitive disadvantage.

It is somewhat ironic that the investment management industry, which has long relied on modern portfolio theory, is only now turning to fund-wide risk measurement systems. These systems have been developed by "quants" on Wall Street who were originally trying to get a grip on their short-term derivatives risk. What we are learning now is that these methods can be extended usefully from the short-term trading environment to the longer-term framework of patient investors.

This turn of events was inevitable. Since advances in technology and communications create almost instantaneous flows of information across the globe, plan sponsors cannot continue to rely on monthly or quarterly hard-copy reports on their investments.

7

Risk Monitoring and Performance Measurement

Learning Objectives

After completing this reading you should be able to:

- Describe the three fundamental dimensions behind risk management and their relation to VaR and tracking error.
- Describe risk planning, including its objectives, effects, and the participants in its development.
- Describe risk budgeting and the role of quantitative methods in risk budgeting.
- Describe risk monitoring and its role in an internal control environment.
- Identify sources of risk consciousness within an organization.
- Describe the objectives and actions of a risk management unit in an investment management firm.
- Describe how risk monitoring can confirm that investment activities are consistent with expectations.
- Describe the Liquidity Duration Statistic and how it can be used to measure liquidity.
- Describe the objectives of performance measurement tools.
- Describe the use of alpha, benchmarks, and peer groups as inputs in performance measurement tools.

Excerpt is Chapter 17 of Modern Investment Management: An Equilibrium Approach, *by Jacob Rosengarten and Peter Zangari.*

00011693-00000002_CH07_p095-116.indd 95 05/07/22 3:09 PM

7.1 OVERVIEW

The *Oxford English Dictionary* describes risk as:

(a) the chance or hazard of commercial loss; also . . .

(b) . . . the chance that is accepted in economic enterprise and considered the source of (an entrepreneur's) profit.

This definition asserts that risk reveals itself in the form of uncertainty. This uncertainty of loss, which risk professionals quantify using the laws of probability, represents the cost that businesses accept to produce profit. Loss potential (i.e., "risk") represents the "shadow price" behind profit expectations. A willingness to accept loss in order to generate profit suggests that a cost benefit process is present. For a return to be deemed desirable, it should attain levels that compensate for the risks incurred.

There are typically policy limits that constrain an organization's willingness to assume risk in order to generate profit. To manage this constraint, many organizations formally budget risk usage through asset allocation policies and methods (e.g., mean-variance optimization techniques). The result yields a blend of assets that will produce a level of expected returns and risk consistent with policy guidelines.

Risk, in financial institutions, is frequently defined as value-at-risk (VaR). VaR refers to the maximum dollar earnings/loss potential associated with a given level of statistical confidence over a given period of time. VaR is alternatively expressed as the number of standard deviations associated with a particular dollar earnings/loss potential over a given period of time. If an asset's returns (or those of an asset class) are normally distributed, 67 percent of all outcomes lie within the asset's average returns plus or minus one standard deviation.

Asset managers use a concept analogous to VaR—called tracking error—to gauge their risk profile relative to a benchmark. In the case of asset managers, clients typically assign a benchmark and a projected risk and return target vis à vis that benchmark for all monies assigned to the asset manager's stewardship. The risk budget is often referred to as tracking error, which is defined as the standard deviation of excess returns (the difference between the portfolio's returns and the benchmark's returns). If excess returns are normally distributed, 67 percent of all outcomes lie within the benchmark's returns plus or minus one standard deviation.

VaR is sometimes expressed as dollar value-at-risk by multiplying the VaR by assets under management. In this manner, the owner of the capital is able to estimate the dollar impact of losses that could be incurred over a given period of time and with a given confidence level. To achieve targeted levels of dollar VaR, owners of capital allocate capital among asset classes (each of which has its own VaR). An owner of capital who wishes to incur only the risks and returns of a particular asset class might invest in an index fund type product that is designed to replicate a particular index with precision. To the extent that the owner wishes to enjoy some discretion around the composition of the index, he or she allows the investment managers to hold views and positions that are somewhat different than the index. The ability to take risks away from the index is often referred to as active management. Tracking error is used to describe the extent to which the investment manager is allowed latitude to differ from the index. For the owner of capital, the VaR associated with any given asset class is based on the combination of the risks associated with the asset class and the risks associated with active management.[1] The same premise holds for the VaR associated with any combination of asset classes and active management related to such asset classes.

By now it is apparent that risk—whether expressed as VaR or tracking error—is a scarce resource in the sense that individuals and organizations place limits on their willingness to accept loss. For any given level of risk assumed, the objective is to engage into as many intelligent profit-making opportunities as possible. If risk is squandered or used unwisely, the ability of the organization to achieve its profit objectives is put at risk. If excessive levels of risk are taken vis à vis budget, the organization is risking unacceptably large losses in order to produce returns that it neither expects nor desires. If too little risk is taken vis à vis budgeted levels, return expectations will likely fall short of budget. The point here is that the ability of an organization to achieve its risk and return targets may be put at risk anytime that risk capital is used wastefully or in amounts inconsistent with the policies established by such organization.

With the above as context, we now delve into the concepts and methods behind risk monitoring and performance measurement in greater depth. The chapter is organized along five themes:

1. We emphasize that risk monitoring is a fundamental part of the internal control environment. It helps ensure that the organization is entering into transactions that are

[1] More formally, the return of the portfolio (R_p) invested in a particular asset class can be described as follows:

$$(R_p) = (R_p - R_a) + R_a$$

where R_a refers to the return of the index or benchmark. The term in parentheses is often referred to as active or excess return. From this expression, one can see that the variance of the portfolio's return (V_p) can be reduced to:

$$V_p = \text{Variance(Excess return)} + \text{Variance(Benchmark)} + 2\,\text{(Covariance between excess return and benchmark return)}$$

The standard deviation of the portfolio is of course the square root of the variance.

authorized and properly scaled; it helps distinguish between events that are unusual and those that should have been anticipated.

2. We show that there are three fundamental dimensions behind risk management—planning, budgeting, and monitoring. We observe that these three dimensions are intimately related and that they can be more completely understood by looking at their commonly used counterparts in the world of financial accounting controls. We posit that there is a direct correspondence between financial planning, financial budgeting, and financial variance monitoring and their risk management counterparts—namely, risk planning, risk budgeting, and risk monitoring.
3. We introduce the concept of a risk management unit (RMU) and describe its role and placement within the organization. We discuss its objectives as well as the need for it to remain independent of portfolio management activities. As we will see, the existence of an independent RMU is a "best practice" for all types of investors, including asset managers, pension funds, and corporations.
4. We describe techniques the RMU uses to monitor exposures in portfolios and provide samples of reports that might be used to deliver such information.
5. Last, we introduce tools that are commonly used in the world of performance measurement. We observe that there is a duality between risk monitoring and performance measurement. Risk monitoring reports on risk that is possible, whereas performance measurement reports on performance (and so risk) that has materialized. We posit that performance measurement is a form of model validation.

We would be remiss if we did not briefly observe that because the sources of risk are many, the modern organization must have a multidisciplinary approach to risk management. In their book, *The Practice of Risk Management*, Robert Litterman and Robert Gumerlock identify at least six distinct sources of risk.[2] These include market, credit, liquidity, settlement, operational, and legal risk. Professional standards, quantitative tools, preemptive actions, internal control systems, and dedicated management teams exist in the modern organization to address each of these. Frequently, these risks overlap and various professional disciplines are required to work together to creatively craft solutions. While in this paper, our primary focus will be management and measurement of market risk and performance, these other risks are ever present and material. Often, stresses in market factors make these other risks more apparent and costly. For this reason, all of these sources of risk are worthy of separate study and investigation.

7.2 THE THREE LEGS OF FINANCIAL ACCOUNTING CONTROL: PLANNING, BUDGETING, AND VARIANCE MONITORING

In the world of financial accounting controls, the concepts of planning, budgeting, and variance monitoring are intimately related. Each is one of the legs of a three-legged stool that defines organizational structure and control. Each leg is fundamental to the success of the organization's raison d'être.

As we will see, the risk management process also can be described as a three-legged stool. Effective risk management processes also have planning, budgeting, and variance monitoring dimensions. It is intuitive that there should exist such a close correspondence between the models that support risk management and those that support financial accounting controls. Remember that risk is the cost of returns—the shadow price of returns. Hence, behind every number in a financial plan or budget there must exist a corresponding risk dimension. This duality suggests that risk management can be described, organized, and implemented using an approach that is already commonly used in the world of financial controls—namely, planning, budgeting, and monitoring.

For a moment, let's focus on the world of financial accounting to explore this point further. Consider how the "financial controls stool" is constructed. The first leg of this stool is a strategic plan or vision that describes earnings targets (e.g., return on equity, earnings per share, etc.) and other goals for the organization (e.g., revenue diversification objectives, geographic location, new product development, market penetration standards, etc.). The strategic plan is a policy statement that broadly articulates bright lines that define points of organizational success or failure.

Once a plan exists, the second leg of the financial controls stool—a financial budget—is created to give form to the plan. The financial budget articulates how assets are to be expended to achieve earnings and other objectives of the plan. The budget represents a financial asset allocation plan that, in the opinion of management, should be followed to best position the organization to achieve the goals laid out in the strategic plan. The budget—a statement of expected revenues and expenses by activity—is a numeric blueprint that quantifies how the strategic plan's broad vision is to be implemented.

[2] *The Practice of Risk Management*, by Robert Litterman and Robert Gumerlock, Euromoney Publications PLC, 1998, p. 32.

The strategic plan and financial budget both presuppose scarcity. In a world of unlimited resources, there is clearly no need for either a budget or a plan. Any mistake could easily be rectified. In a world of scarcity, however, it is apparent that a variance monitoring process—the third leg of the stool—helps ensure that scarce resources are spent wisely in accordance with the guidance offered by the plan and the budget. Monitoring exists because material variances from financial budget put the long-term strategic plan at risk.

In the world of risk management, these same three elements of control—planning, budgeting, and monitoring—apply as well. Although this paper focuses primarily on risk monitoring, it is useful to step back and provide a more complete context for risk monitoring.

7.3 BUILDING THE THREE-LEGGED RISK MANAGEMENT STOOL: THE RISK PLAN, THE RISK BUDGET, AND THE RISK MONITORING PROCESS

The Risk Plan

The following discussion of what constitutes a risk plan may at first blush seem highly theoretical. But upon closer review, the reader will see that sound financial planning standards already incorporate many of the elements that are discussed. We expect many of the ideas referred to here already exist within the body of a comprehensive strategic planning document. For example, most strategic plans include a strengths, weaknesses, opportunities, and threats (SWOT) section in which major risks to the organization are discussed. By introducing the concept of a separate risk plan, however, we are proposing an even greater degree of formality for discussion of risk themes and issues.

We believe that the risk plan should be incorporated as a separate section of the organization's strategic planning document. As such, it should receive all of the vetting and discussion that any other part of the planning document would receive. When in final form, its main themes should be capable of being articulated to analysts, auditors, boards, actuaries, management teams, suppliers of capital, and other interested constituencies.

The risk plan should include five guideposts:

1. The risk plan should set expected return and volatility (e.g., VaR and tracking error) goals for the relevant time period and establish mileposts which would let oversight bodies recognize points of success or failure. The risk plan should use scenario analysis to explore those kinds of factors that could cause the business plan to fail (e.g., identify unaffordable loss scenarios) and strategic responses in the event these factors actually occur. The risk plan helps ensure that responses to events—be they probable or improbable—are planned and not driven by emotion. Difficult business climates have happened before and they will happen again. The planning process should explore the many "paths to the long term" and prepare the organization, and its owners and managers, for the bumps[3] along the way. If any of these bumps are material, concrete contingency plans should be developed and approved by the organization's owners and managers.[4]
2. The risk plan should define points of success or failure. Examples are acceptable levels of return on equity (ROE) or returns on risk capital (RORC). For the purposes of the planning document, risk capital might be defined using value-at-risk (VaR) methods. Since organizations typically report and budget results over various time horizons (monthly, quarterly, annually), separate VaR measures for each time interval should be explored. The VaR (or risk capital) allocated to any activity should be sized in such a way that the exposures and upside associated with the activity are at levels that are deemed appropriate by the organization's owners and managers. A second benefit of attempting to measure the risk capital associated with each activity is that the process helps management understand the uncertainty levels associated with each activity in the plan. The greater the amount of uncertainty and the greater the cost associated with the downside of the VaR estimate actually materializing, the more intensive must be the quality of contingency and remedial planning.
3. The risk plan should paint a vision of how risk capital will be deployed to meet the organization's objectives. For example, the plan should define minimum acceptable RORCs for each allocation of risk capital. In so doing, it helps ensure that the return per unit of risk meets minimum standards for any activity pursued by the organization. The plan should also explore the correlations among each of these RORCs as well to ensure that the consolidated RORC yields an expected ROE, and variability around such expectation, that is at acceptable levels. Finally, the plan should also have a diversification or risk decomposition policy. This policy

[3] In statistical terms, a "bump" might be defined as a three or greater standard deviation event in a relatively short period of time.

[4] Note that scenario analysis can be explored qualitatively as well as quantitatively. In fact, many extreme events lend themselves more to qualitative analysis than quantitative methods.

should address how much of the organization's risk capital should be spent on any one theme.[5]

4. A risk plan helps organizations define the bright line between those events that are merely disappointing and those that inflict serious damage. Strategic responses should exist for any franchise-threatening event—*even if* such events are low-probability situations. The risk plan should identify those types of losses that are so severe that insurance coverage (e.g., asset class puts) should be sought to cover the downside. For example, every organization pays fire insurance premiums to insure against the unaffordable costs of a fire. Fire is one of those events that is so potentially devastating that there is universal agreement on the need to carry insurance protection. Now, consider a more complex example from the world of investment portfolio policy. From an investment standpoint, there may be losses of such magnitude—even if they are infrequent and improbable—that they endanger the long-term viability of the investment plan. For example, firms or plans with large equity holdings[6] could face material loss and earnings variability in the event of protracted and substantial stock market losses. In this case, the risk plan should explore the potential merits of financial insurance (e.g., options on broad market indexes). At a minimum, if such insurance is not purchased, the decision to self-insure should be formally discussed and agreed upon by the organization's owners and management.
5. The risk plan should identify critical dependencies that exist inside and outside the organization. The plan should describe the nature of the responses to be followed if there are breakdowns in such dependencies. Examples of critical dependencies include reliance on key employees and important sources of financing capacity.

   The risk plan should explore how key dependencies behave in good and bad environments.[7] Frequently, very good and or very bad events don't occur in a vacuum; they occur simultaneously with other material events. For example, consider a possible challenge faced by a pension plan. It is conceivable that periods of economic downturn could coincide with lower investment performance, acceleration of liabilities, and a decreased capacity of the contributing organization to fund the plan. For this reason, scenario planning for the pension plan should explore what other factors affect the pension plan's business model in both good and bad environments and develop appropriate steps to help the plan succeed.

An effective risk plan requires the active involvement of the organization's most senior leadership. This involvement creates a mechanism by which risk and return issues are addressed, understood, and articulated to suppliers of capital (owners or beneficiaries), management, and oversight boards. It helps describe the philosophical context for allocations of risk and financial capital and helps organizations ensure that such allocations reflect organizational strengths and underpinnings. It helps organizations discuss and understand the shadow price that must be accepted in order to generate returns.

The existence of a risk plan makes an important statement about how business activities are to be managed. It indicates that owners and managers understand that risk is the fuel that drives returns. It suggests that a higher standard of business maturity is present. Indeed, its very existence demonstrates an understanding that the downside consequences of risk—loss and disappointment—are not unusual. These consequences are directly related to the chance that management and owners accept in seeking profit. This indicates that management aspires to understand the source of profit. The risk plan also promotes an organizational risk awareness and the development of a common language of risk. It demonstrates an intolerance for mistakes/losses that are material, predictable, and avoidable.

The Risk Budget

The risk budget—often called asset allocation—should *quantify* the vision of the plan. Once a plan is put into place, a formal budgeting process should exist to express exactly how risk capital will be allocated such that the organization's strategic vision is likely to be realized. The budget helps the organization stay on course with respect to its risk plan. For each allocation of risk budget, there should be a corresponding (and acceptable) return expectation. For each return expectation, some sense of expected variability around that expectation should be explored. When all of the expected returns, risks, and covariations among risk budgets are considered, the expected return streams, and the variability of such, should be

[5] Diversification policies are routinely included in strategic planning. Such policies take the form of geographic diversification, product diversification, customer base diversification, and so on. Just as organizations produce standards on how much revenue should come from any one source, so too should they examine how much risk originates from any one theme (asset class, portfolio manager, individual security, etc.).

[6] In this context, a "large" holding refers to one that can generate earnings exposures that are deemed material vis à vis the business plan.

[7] Once again, examining correlations among critical business dependencies in periods of stress may be done in a qualitative or quantitative manner.

consistent with the organization's strategic objectives and risk tolerances.

As noted earlier, there are many similarities between financial budgets and risk budgets. Financial budgets calculate net income as the difference between revenue and expenses. ROE is then estimated as net income divided by capital invested. In the case of risk budgets, a risk "charge"—defined as VaR or some other proxy for "risk expense"—can be associated with each line item of projected revenue and expense. Hence, a RORC can be associated with each activity as well as for the aggregation of all activities.

In the case of both financial and risk budgets, presumably ROE and RORC must exceed some minimum levels for them to be deemed acceptable. Both statistics are concerned with whether the organization is sufficiently compensated—in cost/benefit terms—for the expenses and/or risks associated with generating revenues. Just as the financial budget allocates revenue and expense amounts across activities to determine their profitability, so too should a risk budget exist for each activity in order to estimate the *risk-adjusted* profitability of the activity. Just as financial budgets show a contribution to ROE by activity, so too can risk budgets show a contribution to overall risk capital usage by activity. For example, standard mean-variance optimization methods produce estimates of weights to be assigned to each asset class, in addition to overall estimates of portfolio standard deviation and the marginal contribution to risk[8] from each allocation.

Note that both RORC and ROE can and should be estimated over all time intervals that are deemed relevant. For example, if investment boards meet monthly and are likely to react to short-term performance, monthly RORC is relevant. Hence, management must define the time horizons over which risk budget allocations are to be spent and over which RORC should be measured.[9]

An example at this point might be helpful. Assume that an organization has a material investment portfolio. The organization is concerned about the impact of the earnings volatility of this portfolio on reported earnings and, therefore, share price. In constructing a risk budget for this portfolio, the organization might:

- From the risk and business plan, identify acceptable levels of RORC and ROE over various time horizons.
- Using mean variance optimization or other techniques, determine appropriate weights for each investment class.
- Simulate the performance of a portfolio (including the behavior of related liabilities, if relevant) constructed with these weights over various time horizons, and test the sensitivity of this performance to changes in return and covariance assumptions.
- Ensure that the levels of risk assumed at the individual asset class level as well as for the portfolio taken as a whole are at appropriate levels vis à vis the business and risk plan.
- Ensure that the expected variability around expected RORC is at acceptable levels. If there is too much variability vis à vis a competitor's ROE and RORC, the earnings profile might be deemed to be low quality by the marketplace. Accordingly the risk budgeting process must concern itself with not only the absolute magnitude of the RORC at the strategy and overall portfolio levels, but also the variability in such magnitude.
- Explore the downside scenarios associated with each allocation over various time horizons. Ensure that the plan's owners and managers identify such downside as merely disappointing and not unacceptably large (i.e., lethal) given the plan's objectives.
- In each significant downside scenario, loop back to the planning process and ensure that contingency steps exist to bring about a logical and measured response. Ensure that owners, managers, and other outside constituencies (e.g., suppliers of capital) are aware and supportive of these responses.

Clearly, risk budgeting incorporates elements of mathematical modeling. At this point, some readers may assert that quantitative models are prone to failure at the worst possible moments and, as such, are not sufficiently reliable to be used as a control tool. We do not agree. The reality is that budget variances are a fact of life *in both* financial budgeting and risk budgeting. Variances from budget can result from organization-specific factors (e.g., inefficiency) or completely unforeseen anomalies (e.g., macroeconomic events, wars, weather, etc.). Even though such unforeseen events cause ROE variances, some of which may even be large, most managers still find value in the process of financial budgeting. The existence of a variance from budget, per se, is not a reason to condemn the financial budgeting exercise.

[8] The marginal contribution to risk from any asset is defined as the change in risk associated with a small change in the underlying weight of that asset in the portfolio.

[9] We know that risk across different time dimensions does not simply scale by the square root of time. The path to the long term may be much bumpier than a simple scaling might imply. In fact, the long-term result may be entirely consistent with a fair number of short-term anomalies. If so, management must ensure that risk allocations are sized in such a manner that losses associated with short-term market difficulties can be negotiated effectively. Hence, in a manner analogous to financial budgeting, the risk budget helps managers size the bets in each revenue-producing area.

So, too, we believe that the existence of variances from risk budget by unforeseen factors does not mean that the risk budgeting process is irrelevant. To the contrary. Frequently the greatest value of the risk budget derives from the budgeting *process* itself—from the discussions, vetting, arguments, and harmonies that are a natural part of whatever budget is ultimately agreed to. Managers who perform risk budgeting understand that variances from budget are a fact of life and are unavoidable, but are not a reason to avoid a formal risk budgeting process. To the contrary, understanding the causes and extent of such variances and ensuring that appropriate remedial responses exist make the budgeting and planning process even more valuable.

Risk Monitoring

Variance monitoring is a basic financial control tool. Since revenue and expense dollars are scarce, monitoring teams are established to identify material deviations from target. Unusual deviations from target are routinely investigated and explained as part of this process.

If we accept the premise that risk capital is a scarce commodity, it follows that monitoring controls should exist to ensure that risk capital is used in a manner consistent with the risk budget. Material variances from risk budget are threats to the investment vehicle's ability to meet its ROE and RORC targets. If excessive risk is used, unacceptable levels of loss may result. If too little risk is spent, unacceptable shortfalls in earnings may result. Risk monitoring is required to ensure that material deviations from risk budget are detected and addressed in a timely fashion.

7.4 RISK MONITORING—RATIONALE AND ACTIVITIES

There is an increasing sense of risk consciousness among and within organizations. This risk consciousness derives from several sources:

- Banks that lend to investors increasingly care about where assets are placed.
- Boards of investment clients, senior management, investors, and plan sponsors are more knowledgeable of risk matters and have a greater awareness of their oversight responsibilities. Especially as investments become more complicated, there is an increasing focus to ensure that there is effective oversight over asset management activities—whether such activities are managed directly by an organization or delegated to an outside asset manager.
- Investors themselves are expected to have more firsthand knowledge about their investment choices. Perhaps this has been driven, in part, by the notoriety of losses incurred by Procter & Gamble, Unilever, Gibson Greeting Cards, Orange County (California), the Common Fund, and others. After these events, organizations have become interested in stresses and the portfolio's behavior in more unusual environments. Further, in the asset management world, asset managers increasingly must be able to explain, ex ante, how their products will fare in stressful environments. This enhanced client dialogue disclosure is beneficial from two perspectives: First, it raises the level of client confidence in the manager. Second, it reduces the risk of return litigation arising from types of events that were predictable on an ex ante basis.

In response to this heightened level of risk consciousness, many organizations and asset managers have formed independent risk management units (RMUs) that oversee the risk exposures of portfolios and ensure that such exposures are authorized and in line with risk budgets. This trend was definitely spurred on by a highly influential paper authored by the Working Group[10] in 1996.

The Working Group suggested that the RMU's reporting line should incorporate a segregation of duties—a fundamental element of an effective internal controls environment. To be effective, the RMU should be independent in both fact and appearance. This assertion is ratified by industry and professional guidance. For example, the Third Standard produced by the Working Group reads in part:

> *Where possible, an independent internal group . . . should perform oversight. . . . Functions checked independently should include:*
>
> - *Oversight of investment activity*
> - *Limits, monitoring, exception reports and action plans relating to exception reports*
> - *Stress tests and back tests*
> - *. . . Fiduciaries should verify that Managers conduct independent risk oversight of their employees and activities.*

[10] The Working Group was established in April 1996 by 11 individuals from the institutional investment community. Its mission was: "To create a set of risk standards for institutional investment managers and institutional investors." In drafting the final standards, opinions were solicited from a wide range of participants in the financial community including asset managers, academics, plan sponsors, custodians, and regulators. More recently, Paul Myners, in his report (dated March 6, 2001) addressed to the Chancellor of the Exchequer of the United Kingdom entitled *Institutional Investment in the United Kingdom—A Review*, argued persuasively for the increased need for professional development and product understanding of those individuals charged with overseeing pension plans.

In their book, *The Practice of Risk Management*, Robert Gumerlock and Robert Litterman ratify this Standard by stating:

> *It is essential that the risk management function itself must be established independently from the business areas and operate as a controlling or monitoring function. The role of the risk management function is to provide assurance to senior management and the Board that the firm is assessing its risk effectively, and is complying with its own risk management standards. This means that the risk management function has to have an independent reporting line to senior management.*

The risk monitoring unit is a necessary part of the process that ensures best practices and consistency of approach across the firm. It helps ensure that a process exists by which risks are identified, measured, and reported to senior management in a timely fashion. The function is part of an internal control framework designed to safeguard assets and ensure that such assets are managed in accordance with each organization's expectations and management direction.

Objectives of an Independent Risk Management Unit

The objectives of the RMU are:

- The RMU gathers, monitors, analyzes, and distributes risk data to managers, clients, and senior management in order to better understand and control risk. This mission requires that the RMU deliver the right information to the right constituency at the right time.
- The RMU helps the organization develop a disciplined process and framework by which risk topics are identified and addressed. The RMU is part of the process that ensures the adoption and implementation of best risk practices and consistency/comparability of approach and risk consciousness across the firm. As such it is a key promoter of an organization's risk culture and internal control environment.
- To be vibrant, the RMU must be more than a publisher of periodic VaR information. It must also proactively pursue topics and have a topical vein. The RMU should be actively involved in setting and implementing the risk agenda and related initiatives.
- The RMU watches trends in risk as they occur and identifies unusual events to management in a timely fashion. While it is helpful to identify a risk once it is present, it is more meaningful to identify a trend before it becomes a large problem.
- The RMU is a catalyst for a *comprehensive* discussion of risk-related matters, including those matters that do not easily lend themselves to measurement. For example, the RMU should be actively involved in the identification of and organizational response to low-probability yet high-damage events. It should promote discussion throughout the organization and encourage development of a context by which risk data and issues are discussed and internalized.
- The RMU is an element of the risk culture. It should represent one of the nodes of managerial convergence—a locus where risk topics are identified, discussed, and disseminated across the organization and clients. In so doing, it helps promote enhanced risk awareness together with a common risk culture and vocabulary.
- As a part of the internal control environment, the RMU helps ensure that transactions are authorized in accordance with management direction and client expectations. For example, the RMU should measure a portfolio's *potential* (i.e., ex ante) tracking error and ensure that the risk profile is in consonance with expectations.[11]
- Together with portfolio managers and senior management, the RMU identifies and develops risk measurement and performance attribution analytical tools. The RMU also assesses the quality of models used to measure risk. This task involves back testing of models and proactive research into "model risk."
- The RMU develops an inventory of risk data for use in evaluating portfolio managers and market environments. This data, and the methodologies used to create it, must be of a quality and credibility that it is both useful to and accepted by the portfolio managers. This risk data should be synthesized, and routinely circulated to the appropriate decision makers and members of senior management.
- The RMU provides tools for both senior management and individual portfolio management to better understand risk in individual portfolios and the source of performance. It establishes risk reporting and performance attribution systems to portfolio managers and senior management. In the process, the RMU promotes transparency of risk information.
- The RMU should not manage risk, which is the responsibility of the individual portfolio managers, but rather *measure* risk for use by those with a vested interest in the process. The RMU cannot reduce or replace the decision methods and

[11] For asset management firms, this oversight spans a different dimension of risk than the function currently performed by compliance departments. In fact, the RMU forms a natural complement to the efforts of the compliance department within asset management firms. By definition, the matching of actual positions with guidelines by the compliance department involves examining events that have already happened. In contrast, by stressing data and exploring both common and uncommon scenarios, the RMU explores the implications of what *might* happen in the future.

responsibilities of portfolio managers. It also cannot replace the activities of quantitative and risk support professionals currently working for the portfolio managers. Trading decisions and the related software and research that support these decisions should remain the responsibility of the portfolio managers and their support staffs. The RMU measures the extent to which portfolio managers trade in consonance with product objectives, management expectations, and client mandates. If the RMU finds what it deems to be unusual activities or risk profiles, it should be charged with bringing these to the attention of the portfolio managers and senior management so that an appropriate response can be developed and implemented.

Examples of the Risk Management Unit in Action

An effective internal control environment requires timely, meaningful, and accurate information flows between senior management and the rest of the organization. Information flows allow management to ask questions. Questions and the ability to probe into the process by which the business operates are fundamental to loss avoidance and profit maximization.

Risk monitoring is principally concerned with whether investment activities are behaving as expected. This suggests that there should be clear direction as to what results and risk profiles should be deemed normal versus abnormal. It is our experience that the very best managers in the world achieve success in no small part because they have a time-tested conviction and a philosophy that has a stable footprint. For example, the best growth managers do not invest in value themes; the best U.S. fixed income managers do not take most of their risk in non-U.S. instruments; and so on. In fact, the premier managers remain true to their time-tested convictions, styles, and philosophies. Further, the best managers apply well-defined limits—expressed both in absolute terms as well as in marginal contribution to risk terms—on how they spend any given amount of risk budget. The result of this discipline is a portfolio that produces a return distribution that meets the following world-class standards:

- It is consistent with client expectations. The risk capital consumed by the manager approximates the amount of risk budget the client authorized the manager to spend.
- It is derived from organizational or individual strengths (e.g., stock selection, sectors, of the market growth or value, portfolio construction techniques, etc.).
- It is high-quality in the sense that it is not the result of luck, but rather of sound organizational plans and decisions that have been executed in accordance with philosophy and conviction.
- It is the result of a well-articulated and well-defined process and risk culture whose major elements are understood and embodied by the organization.
- It is stable, consistent, and controlled. It produces results that can be explained and repeated across time with a high degree of confidence.

The RMU helps create systems to report risk information to interested constituencies (senior management, control nodes, portfolio managers, etc.). This information should reveal several broad themes. In particular, it should allow the user to be conclusive concerning:

- Whether the manager is generating a forecasted level of tracking error that is consistent with the target established by the mandate.
- Whether, for each portfolio taken, individually and for the sum of all portfolios taken as a whole, risk capital is spent in the expected themes.
- Whether the risk forecasting model is behaving as predicted.

Is the Forecasted Tracking Error Consistent with the Target?

The forecasted tracking error is an estimate of the potential risk that can be inferred from the positions held by the portfolio derived from statistical or other forward-looking estimation techniques. An effective risk process requires that portfolio managers take an appropriate level of risk (i.e., neither too high nor too low) vis à vis client expectations. This forecast should be run for each individual portfolio as well as for the sum of all portfolios owned by the client. Tracking error forecasts should be compared to tracking error budgets[12] for reasonableness. Policy standards should determine what magnitude of variance from target should be deemed so unusual as to prompt a question and what magnitude is so material as to prompt immediate corrective action. In this manner, unusual deviations across accounts will be easier to identify.

Figure 7.1 is an example of a tracking error forecast report for a sample U.S. equity fund produced by Goldman Sachs Asset Management (GSAM) on its proprietary portfolio analysis and construction environment (PACE) platform. PACE is a risk and return attribution system that we use to forecast risk across the spectrum of equities managed by GSAM. Observe from the header of this report that the forecasted tracking error for this account, as estimated by the PACE model, is 3.68 percent per annum. A second equity factor risk model, Barra, projects a tracking error forecast of 2.57 percent. Since each model

[12] Tracking error budgets should exist for each portfolio and be determined as part of the organization's asset allocation process.

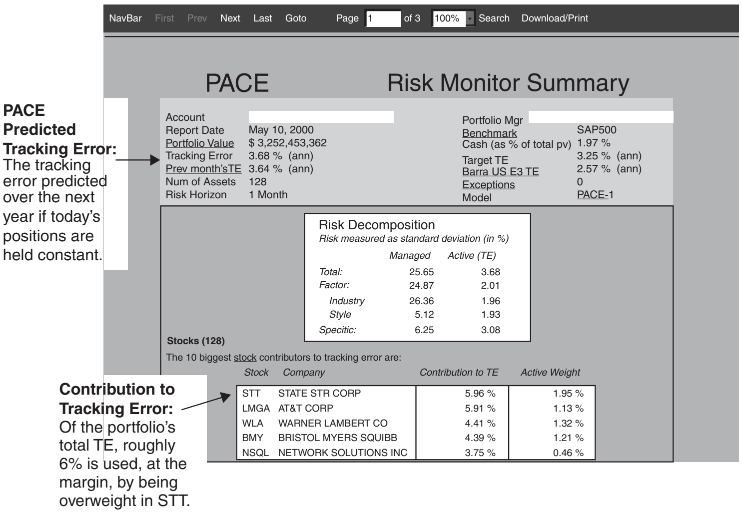


Figure 7.1 Risk report for a U.S. equity fund.

uses different assumptions to forecast risk, it is not surprising that two different models would produce different results. What is comforting in this case is that both measures of risk are comparable to the targeted risk level of 3.25 percent per annum.

This same report should be produced for each account that is supposed to be managed in a parallel manner to ensure consistency of overall risk levels.

Is Risk Capital Spent in the Expected Themes for Each Portfolio?

In financial variance monitoring, it is insufficient to know *only* that the overall expense levels are in line with expectations. Each line item that makes up the total must also correspond to expectations. If there are material variances among line items that tend to offset each other, the person monitoring variances should be on notice that unusual activity may be present. As an example, if a department meets its overall expense budget but is materially over budget in legal fees (with favorable offsets in other areas), the reviewer might conclude that an event is present that might put future returns at risk.

The same principle holds for risk monitoring. Managers should be able not only to articulate overall tracking error expectations, but also to identify how such tracking error is decomposed into its constituent parts. This will let the risk manager opine on whether risk is being incurred in accordance with expectations both in total as well as at the constituent level. If the risk decomposition is not in keeping with expectations, the manager may not be investing in accordance with the stated philosophy.

This type of situation is often referred to as "style drift." An example of this might be a growth manager who is investing in consonance with the correct overall tracking error target, but who is placing most of the risk in value themes. In this case, the investor is acquiring the correct level of overall risk, but the wrong style decomposition.

Examples of risk decomposition that a manager should be able to articulate and which the RMU should monitor might include:

- The range of acceptable active weights (portfolio holdings less benchmark holdings) at the stock, industry, sector, and country levels.
- The range of acceptable marginal contributions to risk at the stock, industry, sector, and country levels.

Refer again to Figure 7.1. For this particular portfolio, we observe that State Street Corp. represents an active weight of 1.95 percent of the total portfolio and that its marginal contribution to tracking error is 5.96 percent. The risk monitoring function should conclude as to whether this active weight and risk decomposition—which may alternatively be described as the portfolio's diversification footprint—is in line with expectations. What is being measured here is the extent to which the manager is investing capital in accordance with stated policies. This report should be run at the manager level as well as at the consolidated portfolio level to ensure that no undue (i.e., unacceptably large vis à vis budget) concentrations of risk are present.

Figure 7.2 shows the largest active exposures and marginal contributions at the industry level. The risk monitor should be able to opine on whether the levels of risk concentration observed are in accordance with manager philosophy. Once again, this report should be run at the manager level as well as at the consolidated portfolio level to ensure that no undue (i.e., unacceptably large vis à vis budget) concentrations of risk are present that might put either a strategy or the overall plan at risk.

Is the Risk Forecasting Model Behaving as Predicted?

As indicated earlier, the risk forecasting model uses statistical methods to produce a forward-looking estimate of tracking error. Accordingly, the risk monitor is charged with knowing whether the model is producing meaningful estimates of risk.

For example, GSAM's PACE tabulates the number of times that a portfolio's actual return is materially different from its risk forecasts. As an example of this test, please refer to Figure 7.3.

that actually occurred. To examine these implied paths, Monte Carlo methods are commonly applied.

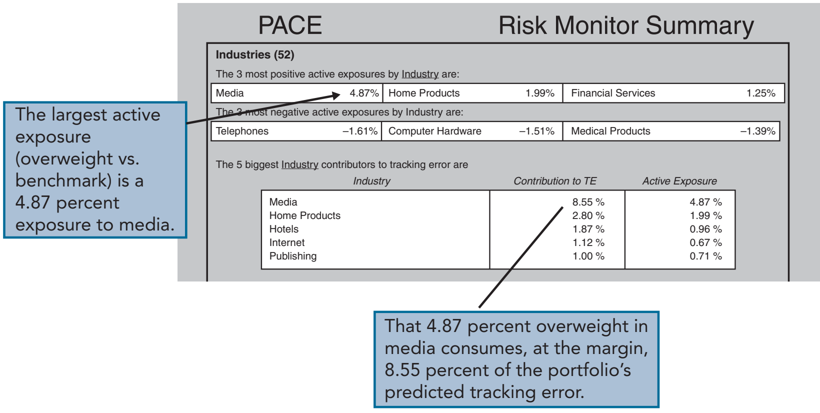


Figure 7.2 **Industry-level exposures and marginal contributions.**

Figure 7.4 graphs the results of a Monte Carlo simulation for a sample equity portfolio that was prepared to study how tracking error forecasts fluctuate depending on the environment used to estimate the risk forecast.[14] Note that as of April 26, 2002, for this portfolio, the PACE risk model projected a tracking error of 5.08 percent per annum. The tracking error target for this portfolio was 5 percent. So, at first blush, it seems as though the portfolio has an overall risk profile that is closely aligned with the risk target. Common sense tells us, however, that the particular combination of assets held in the portfolio might exhibit quite different tracking error characteristics in different environments.

Note that if the model is behaving as expected, the portfolio's actual returns should exceed the tracking error forecast by approximately one day per month. Over the four months ended April 30, one therefore expects that there should be four occurrences where actual returns exceed forecast. In fact, there are three. The risk monitor can conclude that the model is behaving appropriately over the period. Had this result not been reached, some of the model's assumptions might have needed to be revisited.

Note from Figure 7.3 that this technique gives no guidance as to how much the model might underestimate risk in the event that the actual result exceeds forecast. It only explores the frequency with which this result occurs. The risk monitoring professional should also explore how tracking error might behave in more unusual circumstances.[13]

There are many ways to examine how a portfolio might behave during periods of stress. One technique is historical simulation. To apply this approach, one takes today's positions and applies historical price changes to them to see what the earnings impact would have been had such positions been held fixed over a period of time. A shortfall of this method is that observed history produces only one set of realized outcomes. A more robust approach would allow us to examine the myriad outcomes that are probabilistically implied by the one set of outcomes

The PACE forecast is derived by assuming that the underlying data have a halflife of about half a year. When estimating the covariance matrix[15] that is at the heart of the risk forecast, data that are six months old are weighted half as much as current data, and data that are one year old are weighted about one-quarter of current data, and so on. So, more import is given to recent data than to aged data in forecasting risk. This key assumption means that the co-variance matrix itself fluctuates over time not only because different data are used to estimate its components but also because the passage of time causes the import of any particular element in the matrix to have an ever smaller weight.

To examine how a tracking error forecast might fluctuate over time, Figure 7.4 simulates the frequency distribution of the tracking error of the positions held at April 26, 2002, over the period from June 1998 until April 26, 2002. These positions, when introduced into the Monte Carlo engine, would have yielded an average tracking error forecast that would have peaked at 6.5 percent in late 1998 and mid-2000. At these times, the 98th percentile risk forecast reached levels of 7 percent.

[13] It is often true that a three standard deviation scenario is more draconian than that value that is implied by multiplying a one standard deviation loss by three. This result occurs for two reasons: (1) Many products have nonlinear payoff structures (i.e., embedded options); and (2) the global stresses that are present in a three standard deviation scenario are qualitatively different than those which are present in a one standard deviation scenario. As an example, counterparty credit risk increases in more unusual environments.

[14] It is beyond the scope of this chapter to delve in depth into the calculation methodology behind Monte Carlo methods. Rather we present an output of a Monte Carlo analysis to give the reader a sense as to the types of insights it might provide.

[15] Recall that the standard deviation (or tracking error) is calculated by the formula: Tracking error $= [W^T \Sigma W]^{1/2}$ where W is an $N \times 1$ matrix of weights applied to particular factors (e.g., risk factors, or market value of stock holdings, etc.) and Σ represents the $N \times N$ covariance matrix associated with the returns of these factors.

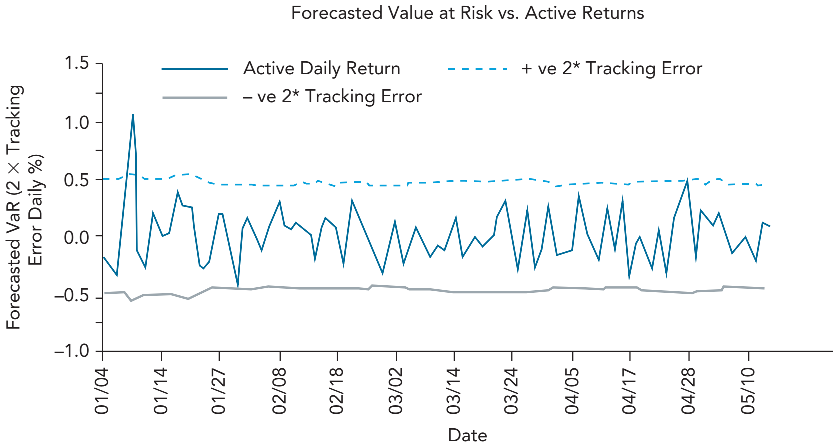


Figure 7.3 **Model validation.**

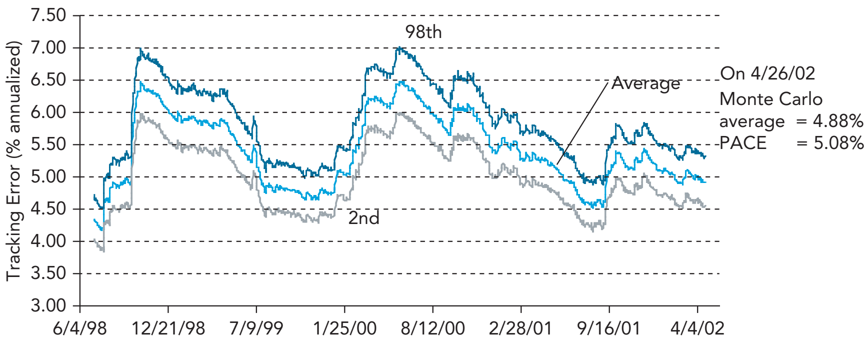


Figure 7.4 **Example of Monte Carlo methods to explore tracking error stresses.**

The risk monitoring professional should consider whether these ranges of tracking error that might occur during periods of stress fall within acceptable levels vis à vis the long-term target of 5 percent. If these levels of tracking error are deemed unacceptably large, an appropriate response might be to run the portfolio at a lower risk profile (say, 4 percent) such that there is reason to believe that the tracking error is less likely to reach unacceptably large levels during periods of stress.[16]

Quantifying Illiquidity Concerns

Since a portfolio's liquidity profile can change dramatically during difficult market environments, tools that measure portfolio liquidity are an essential element of the stress analysis. For example, investors must be aware if a partial redemption could cause an illiquid asset to exceed some guideline.[17] Since redemption risk can correlate with difficult markets, some illiquid situations (e.g., 144A securities, position concentrations, etc.) can coincide with unanticipated redemptions of capital.[18] The risks associated with many of these situations are often apparent only if large stresses are assumed. A tool we use at GSAM to assess the potential implications of illiquidity is the "liquidity duration" statistic.

To calculate this statistic, begin by estimating the average number of days required to liquidate a portfolio assuming that the firm does not wish to exceed a specified percent of the daily volume in any given security. The point here is that we wish to estimate how long it would take to liquidate a portfolio's holdings in an orderly fashion—that is, without material market impact. For example, suppose that we do not wish to exceed more than 15 percent of the daily volume in any given security holding. The number of days required to liquidate any given security we term the liquidity duration for that security. More precisely, the liquidity duration for security i can be defined as:

$$LD_i = Q_i/(.15 \cdot V_i)$$

where LD_i = Liquidity duration statistic for security i, assuming that we do not wish to exceed 15% of the daily volume in that security

Q_i = Number of shares held in security i

V_i = Daily volume of security i

An estimate of liquidity duration for the portfolio taken as a whole can be derived by weighting each security's liquidity duration by that security's weight in the portfolio.

[16] Recall that tracking error is shorthand for the magnitude of earnings variability associated with a certain degree of statistical confidence. If this variability is unacceptably large, it may place the organization's overall strategic plan and goals at risk.

[17] As an example, a U.S. mutual fund cannot hold more than 15 percent of its assets in illiquid securities.

[18] An example of this statement is the acceleration of liabilities in a pension plan due to increases in early retirements in periods of recession.

Liquidity duration is readily calculated for equity holdings, as volume data are easily available. In the case of fixed income securities, where volume information is not available, the estimate of the number of days required to liquidate a position—and an overall portfolio—in an orderly fashion (i.e., without a material adverse earnings impact) will likely result from discussions with portfolio managers.

Credit Risk Monitoring

For the purposes of this discussion, we assume that the credit risk of each instrument is researched and understood by the portfolio manager. We further assume that through factor models or other techniques, the RMU professional can estimate the VaR or tracking error consequences of credit exposures imbedded in the securities held by the portfolio.

In addition to quantifying security-specific and overall portfolio credit exposure, it is important that the RMU understand the credit consequences of dealing with brokers, custodians, execution counterparties, and the like. It is a truism that credit risk is frequently the other side of the coin of market risk. Discussions on market risk are often, at their heart, driven by credit matters. In certain asset classes (e.g., emerging markets) credit risk and market risk may be virtually inseparable. Further, since credit risk is an attribute of performance, it should also be an element of the risk process. As an example, many global indexes (e.g., IFC) now include emerging market countries. To the extent that financial systems in such countries (e.g., Egypt and Russia) are evolving and immature, institutions face credit risk when settling trades. The expected return on such transactions is a function not only of issuer-specific risk, but of credit/settlement risk as well. For this reason, the RMU should ensure that all counterparties used to execute and settle trades meet credit policy criteria.

7.5 PERFORMANCE MEASUREMENT—TOOLS AND THEORY

Until now, we have largely focused our attention on measuring potential risk—an estimate of the risk and return that is possible. The other side of this coin is measurement of realized outcomes. In theory, if the ex ante forecasts are meaningful, they should be validated by the actual outcomes experienced. In this sense, performance measurement might be thought of as a form of risk model validation.

In general, the objectives of performance measurement tools are:

- To determine whether a manager generates consistent excess risk-adjusted performance vis à vis a benchmark.
- To determine whether a manager generates superior risk-adjusted performance vis à vis the peer group.
- To determine whether the returns achieved are sufficient to compensate for the risk assumed in cost/benefit terms.
- To provide a basis for identifying those managers whose processes generate high-quality excess risk-adjusted returns. We believe that consistently superior risk-adjusted performance results suggest that a manager's processes, and the resulting performance, can be replicated in the future, making the returns high-quality.

Reasons That Support Using Multiple Performance Measurement Tools

To calculate a risk-adjusted performance measure, two items must be known:

1. Returns over the relevant time period.[19]
2. Risk incurred to achieve such returns.

Risk is ultimately a very human concept comprised of many human dimensions (e.g., emotion, psychological response to uncertainty, fear of underperformance, etc.). Since no two human beings are identical, no two risk assessments are identical. To measure risk and return most comprehensively, we have seen that a panoply of tools (e.g., historical simulations, liquidity awareness, Monte Carlo methods, etc.) can be helpful in order to gain the most complete understanding of the risk present in a portfolio. If the tools yield materially different forecasts, the onus is on the risk professional, working together with senior management and portfolio managers, to apply judgment to determine the most appropriate forecast under the circumstances.

How to Improve the Meaningfulness of Performance Measurement Tools

Performance tools are especially robust when they confirm a priori expectations regarding the quality of returns. If we can identify a disciplined and effective process, we should expect that the process will generate superior risk-adjusted returns. The tools provide a means of measuring the extent of the process's effectiveness. The tools should confirm our

[19] In cases where a portfolio holds illiquid assets, returns are the product of human judgment to some degree. It is conceivable that two individuals looking at the same positions could arrive at materially different valuations—this phenomenon occurs because there can be a material divergence between value and price in illiquid markets. In contrast, for liquid securities, the low bid/ask spread is an indication that price is a good approximation of value.

belief that the process is indeed functioning the way it was designed to. For example, risk decomposition analysis should show that small cap managers are in fact taking most of their risk in small cap themes. Similarly, a manager with a particular industry specialization should be able to demonstrate that most of that risk budget is spent in securities in that industry. And so on.

For a process to be present, one must be able to define "normal behavior." If normalcy is not identified, the process is likely to be too amorphous to be quantified. Simply put, a process cannot exist without well-defined expectations and decision rules.

Normal behavior suggests that behavior should be predictable. If a process is effective, continued normal behavior (i.e., trading in a manner consistent with the established process) should give us reason to conclude that high-quality returns observed in the past are likely to replicate themselves in the future.

Later on in this chapter, we will introduce some commonly used performance tools. Before discussing these, however, it is worth noting that performance tools, while necessary, are not a substitute for timely management intervention when there is an indication of abnormal behavior. By the time that abnormal behavior manifests itself in the form of poor performance statistics, the damage might already be irreversible. For this reason, we believe that performance tools must be supplemented with:

- A clear articulation of management philosophy from each portfolio manager. This philosophy statement should identify how the manager expects to extract returns from the market. It should identify ways of knowing when the manager's process is successful and when it is unsuccessful.
- A routine position and style monitoring process designed to identify deviations from philosophy or process. This is a type of early warning system.

Appendix A at the end of this chapter gives examples of the kinds of information that might be obtained from each manager to help the RMU define and understand each manager's investment philosophy more completely. This list is not meant to be exhaustive, nor is it appropriate for every organization and manager. We provide it here as an example of techniques used in identifying and monitoring "normalcy."

For quantitative portfolio measurement tools to be effective, we must have a sufficient number of data points to form a conclusion with a certain level of statistical confidence. For the purposes of the remainder of this chapter, we will assume away this issue. In practice, however, the dearth of performance data often hinders the effectiveness of performance measurement tools. In such cases, the organization will be even more dependent on measuring compliance with manager philosophy.[20]

At this point, we turn our focus to identifying some commonly used performance tools and techniques. (Appendix B, for the reader's reference, is a more mathematical treatment of performance calculation methodologies.)

Tool #1—The Green Zone

Each portfolio manager should be evaluated not only on the basis of ability to produce a portfolio with potential (i.e., forecasted) risk characteristics comparable to target, but also on the basis of being able to *achieve* actual risk levels that approximate target. A manager who can accomplish this task, and earn excess returns in the process, has demonstrated the ability to anticipate, react to, and profit from changing economic circumstances.

At GSAM, we have pioneered a concept called the green zone[21] to identify instances of performance or achieved tracking error that are outside of normal expectations. The green zone concept embodies the following elements:

1. For the prior week, month, and rolling 12 months, we calculate the portfolio's *normalized returns*, which are defined as excess returns over the period minus budgeted excess returns over such period, all divided by target tracking error scaled for time.[22] This statistic might be viewed as a test of the null hypothesis that the achieved levels of excess

[20] Even though an organization lacks sufficient data to measure the effectiveness of many managers based on their historical results, it still has sufficient information to conclude whether:

- A manager's philosophy and practices meet commonsense criteria and are likely to extract risk-adjusted performance from the market.
- Each manager's portfolio is consistent with stated philosophy. For example, the RMU should be able to determine that the current portfolio has overall risk levels and risk decomposition characteristics that conform to the manager's philosophy.

An administrative process that measures congruence between manager philosophy and actual trades, money management behavior, loss control, position sizing, and so on is also a form of performance measurement, although not one that we intend to deal with in this paper. If the manager cannot articulate his portfolio management techniques effectively, and if adherence to stated techniques cannot be measured, it is difficult to conclude that a process exists which can be replicated successfully in the future.

[21] Refer to an article entitled: "The Green Zone . . . Assessing the Quality of Returns," by Robert Litterman, Jacques Longerstaey, Jacob Rosengarten, and Kurt Winkelmann of Goldman Sachs & Co. (March 2000).

[22] For example, in calculating the monthly normalized return, the denominator consists of the annual tracking error target divided by the square root of 12.

00011693-00000002_CH07_p095-116.indd 108 05/07/22 3:10 PM

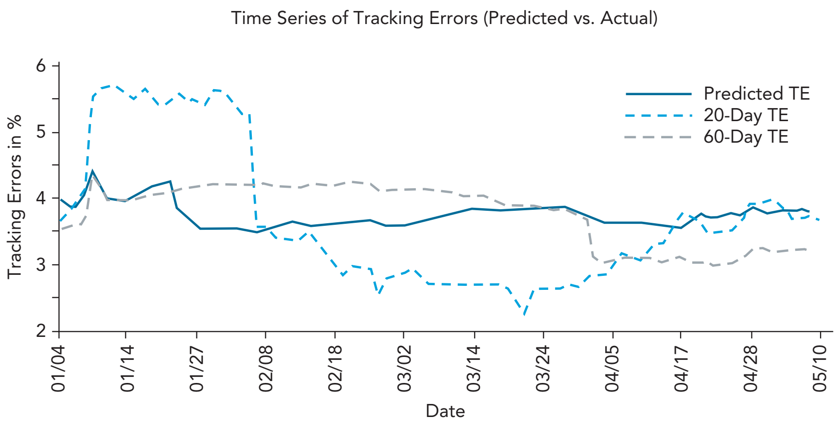


Figure 7.5 **Example of rolling 20- and 60-day tracking errors (annualized).**

returns are statistically different from the targeted/budgeted excess returns.

2. For the prior 20- and 60-day periods, we calculate the ratio of annualized tracking error to targeted tracking error. In this test, we examine whether the variability in excess returns is statistically comparable to what was expected.[23] Note that there is no one correct period of time over which to measure tracking error. While for the purposes of this chapter we have selected a shorter-term horizon, strong arguments can be made for including longer-term horizons as well. The point here is that unusual blips in volatility may serve as filters for identifying anomalous environments in which underlying risk dimensions may be undergoing profound change. This tool is designed to help management and portfolio managers ask better and timelier questions.

   As an example of this point, consider Figure 7.5, which shows the time series of predicted tracking errors juxtaposed against rolling 20- and 60-day tracking errors. Not surprisingly, the 20-day measure is more volatile than the 60-day measure and is therefore more responsive to changes in market behavior. The challenge for the risk monitoring professional is to ascertain whether the signal is anomalous or whether it carries information content that should be acted upon. At GSAM, we use this signal as a basis for initiating dialogue between the RMU and portfolio managers to better understand the causes behind these two signals and their consequences.

3. For each of the calculations in (1) and (2) above, we form policy decisions about what type of deviation from expectation is large enough, from a statistical standpoint, to say that it does not fall in the zone of reasonable expectations that we call the green zone. If an event is unusual, but still is expected to occur with some regularity, we term it a yellow zone event. Finally, red zone events are defined as truly unusual and requiring immediate follow-up. The definition of when one zone ends and a second begins is a policy consideration that is a function of how certain we would like to be that all truly unusual events are detected in a timely fashion. For example, if the cost of an unusual event is very high, one would expect a very narrow green zone and quite wide yellow and red zones. In this case, one would expect to find more false positives, which are by-products of the policy's conservatism.

4. The results of the green zone analysis are summarized in a document of the form shown in Figure 7.6. What follows is a brief description of this document excerpted from the article entitled *The Green Zone . . . Assessing the Quality of Returns*.

   [In Figure 7.6] we show an example of a portion of one of our weekly performance reports (using hypothetical products). This report, known internally as the "green sheet," has columns that are color-coded for easy recognition of signals of tracking error concerns. For example, we have defined the green zone for a hypothetical set of U.S. equity portfolios, including all ratios of realized 20-day tracking error to target between .7 and 1.4, and have defined the red zone as ratios below .6 or above 2. For the 60-day tracking error we define the green zone as the range between .8 and 1.3. The red zone is defined as ratios below .7 or above 1.8.

   . . . the predefined green, yellow, and red zones provide clear expectations for the asset management division portfolio managers. When portfolios move into the yellow or red zone, which will happen every so often, it may be time for a discussion of what is going on. We never expect portfolio management, or risk monitoring, to be reduced to a formula, but these types of quantitative tools have proved to be useful in setting expectations and in providing useful feedback which can foster better quality control of the investment management process.

[23] This test is analogous to ANOVA techniques (e.g., the "F" test) in which one looks at the ratio of variances to determine whether they are statistically comparable. In this case, we are examining the ratio of standard deviations.

Net Performance		Normalized Return vs. Target					Annualized Tracking Error										
							Last	Last	Last20D	Last60D	Target	Prior Week			Month To Date		
		Week	MTD	YTD	3Mo	12Mo	20D	60D	/Target	/Target	TE	P	B	D	P	B	
Portfolio	Benchmark																
U.S. Equity Portfolios																	
Portfolio 1	S&P 500	–1.21	–0.22	–0.16	–0.19	0.45	263	260	1.05	1.04	250	0.76	1.18	–0.42	–1.02	–0.91	–(
Portfolio 2	R 1000 Growth	0.37	0.91	1.70	3.14	0.91	526	390	1.75	1.30	300	2.13	1.98	0.15	0.62	0.06	(
Portfolio 3	R 1000 Value	0.26	0.85	1.44	1.22	1.36	226	240	0.90	0.96	250	1.32	1.23	0.09	–0.39	–0.83	(
Portfolio 4	R 2000	0.28	–0.12	0.37	–0.07	–0.68	300	288	0.86	0.82	350	4.39	4.25	0.14	1.68	1.76	–(

Figure 7.6 Representative green sheet.

Note: This chart is to be used for illustrative purposes only. These are not, and should not be viewed as, predictions or projections of future returns of those classes of assets or any investments, including any fund or separate account managed by GSAM, Goldman Sachs & Co., or any other brokerage account.

The point here is that portfolio managers will likely find most meaningful those techniques that measure and describe risk in the same manner that they internalize these issues. Once again, this argues for having a range of risk and attribution models in order to achieve the most robust understanding.

Tool #2—Attribution of Returns

A commonly used tool to measure the quality of returns is performance attribution. This technique attributes the source of returns to individual securities and/or common factors. Recall that when analyzing the risk profile of a portfolio, we discussed techniques (e.g., risk decomposition) to measure the extent to which the implied risks in a portfolio are consistent with expectations and manager philosophy. So, too, when examining the actual returns of a portfolio, we are concerned that the returns were sourced from those themes where the manager intended to take risk and that such returns are consistent with the risks implied by the *ex ante* risk analysis.

One form of attribution, commonly called variance analysis, shows the contribution to overall performance for each security in the portfolio. Figure 7.7 is an excerpt of this kind of analysis for a stock portfolio. This same kind of analysis can be performed at the industry, sector, and country levels, essentially by combining the performance of individual securities into the correct groupings. The RMU professional can use this analysis to ascertain whether the portfolio tended to earn returns in those securities, industries, sectors, and countries where the risk model indicated that the risk budget was being spent.

To the extent that the manager thinks of risk in factor space as opposed to security-specific space, the attribution process can be performed on this basis. Namely, the attribution process captures the weightings in various risk factors on a periodic basis and also accumulates the returns to such factors in order to produce a variance analysis expressed in factor terms.

As a general rule, it is most meaningful to attribute returns on the same basis that *ex ante* risk for such returns is measured. For managers who think in factor terms, factor risk analysis and factor attribution will likely be more meaningful. For managers who think about risk in terms of individual securities, risk forecasting and attribution at the security level will likely be more relevant. This is not to say that risk should not be measured using a range of models.

Tool #3—The Sharpe and Information Ratios

The Sharpe ratio divides a portfolio's return in excess of the risk-free rate by the portfolio's standard deviation. The information ratio divides a portfolio's excess returns (vis à vis the benchmark) by the portfolio's tracking error. Both of these tools are designed to produce estimates of risk-adjusted returns, where risk is defined in standard deviation or tracking error space.

In theory, two different estimates of standard deviation (or tracking error) could be used for these ratios—actual levels of standard deviation as well as forecasted levels. In our judgment, both are relevant. There are occasions where the realized risk—the risk actually observed by the investor—is materially different from the potential risk forecasted by a risk model.[24] In the Monte Carlo analysis in Figure 7.4 we saw how stress tests can be used to provide a picture of how identical holdings can have quite different return and risk characteristics depending on the environment. If the estimates of potential risk capture these stressed scenarios, potential risk might well exceed realized risk. A favorable Sharpe or information ratio calculated using *realized* risk might be much less attractive when expressed in *potential* risk space. Over time, if the risk model is accurate, the realized risk will center on the potential risk.

The Sharpe and information ratios incorporate the following strengths:

- They can be used to measure relative performance vis à vis the competition by identifying managers who generate superior risk-adjusted excess returns vis à vis a relevant peer group. RMUs and investors might specify some minimum rate

[24] Risk models attempt to measure potential risk. Ultimately, the true potential risk is not knowable. We only see its footprints over time in the form of realized risk. Still, even this realized risk is only one outcome of an infinite number of outcomes that were in theory possible.

PACE **Multi Period Attribution**

Contributors to Active Return by Asset ***(All entries in %)***

Top 30 *(Ranked from highest to lowest Active Contrib.)*

Stock	Company	Contribution	Avg. Active Wgt.
INTC	INTEL CORP	1.45	–3.70
INTU	INTUIT INC	0.88	4.97
JNJ	JOHNSON & JOHNSON	0.46	–3.05
HD	HOME DEPOT INC	0.41	–2.11
GE	GENERAL ELECTRIC CO	0.34	–6.73
CSCO	CISCO SYSTEMS INC	0.27	–2.50
ABT	ABBOTT LABS	0.25	–0.89
IBM	INTERNATIONAL BUSINESS	0.25	–2.06
ENR	ENERGIZER HOLDINGS INC	0.23	2.20
MO	PHILIP MORRIS COS INC	0.23	–1.08
TXN	TEXAS INSTRUMENTS INC	0.23	–1.06
AMAT	APPLIED MATERIALS INC	0.21	–0.81
OMC	OMNICOM GROUP INC	0.17	–0.30
MRK	MERCK & CO INC	0.16	–1.69
WYE	WYETH	0.16	–1.60

Bottom 30 *(Ranked from lowest to highest Active Contrib.)*

Stock	Company	Contribution	Avg. Active Wgt.
CVC	CABLEVISION SYSTEMS CORP	–0.86	1.30
CCU	CLEAR CHANNEL	–0.81	2.45
UVN	UNIVISION COMMUNICATIONS	–0.78	3.78
RMG	RAINBOW MEDIA GROUP	–0.66	0.88
VIA.B	VIACOM INC	–0.61	8.59
DISH	ECHOSTAR COMMUNICATIONS	–0.60	2.52
HET	HARRAHS ENTERTAINMENT	–0.56	7.65
EVC	ENTRAVISION	–0.47	2.03
MGM	METRO GOLDWYN MAYER INC	–0.45	1.68
WON	WESTWOOD ONE INC	–0.41	3.88
FRE	FEDERAL HOME LOAN	–0.41	5.32
FNM	FEDERAL NATIONAL	–0.40	5.78
PCS	SPRINT CORP	–0.35	0.51
CD	CENDANT CORP	–0.34	3.03
TSG	SABRE GROUP HOLDINGS INC	–0.33	3.37

Figure 7.7 **Sample variance analysis.**

of acceptable risk-adjusted return when evaluating manager performance.

- They test whether the manager has generated sufficient excess returns to compensate for the risk assumed.
- The statistics can be applied both at the portfolio level as well as for individual industrial sectors and countries. For example, they can help determine which managers have excess risk-adjusted performance at the sector or country level.

The Sharpe and information ratios incorporate the following weaknesses:

- They may require data that may not be available for either the manager or many of his competitors. Often an insufficient history is present for one to be conclusive about the attractiveness of the risk-adjusted returns.
- When one calculates the statistic based on achieved risk instead of potential risk, the statistic's relevance depends, to some degree, on whether the environment is friendly to the manager.

Tool #4—Alpha versus the Benchmark

This tool regresses the excess returns of the fund against the excess returns of the benchmark.

The outputs of this regression are:

- An intercept, often referred to as "alpha," or skill.
- A slope coefficient against the excess returns of the benchmark, often referred to as "beta."

Standard confidence tests can be applied to the regression's outputs. The alpha term can be tested for statistical significance to see if it is both positive and statistically different from zero.

This performance tool incorporates the following strengths:

- It allows management to opine whether skill is truly present or excess returns are happenstance. It tests whether the manager has generated excess returns vis à vis the benchmark.
- It allows management to distinguish between excess returns due to leverage and excess returns due to skill.
- The alpha and beta statistics, and tests of significance, are easy to calculate.
- The beta statistic shows if an element of the manager's returns are derived from being overweight or underweight the market (occurs if the beta is statistically different from 1.0).

This performance tool incorporates the following weakness:

- There may not be a sufficient number of data points to permit a satisfactory conclusion about the statistical significance of alpha.

Tool #5—Alpha versus the Peer Group

This tool regresses the manager's excess returns against the excess returns of the manager's peer group. It is used to determine whether the manager demonstrates skill over and above what is found in the peer group.

The peer group's return is the capital-weighted average return of all managers who trade comparable strategies. The peer group is basically the manager's competitors in his strategy.

The outputs of this regression are:

- An intercept, often referred to as "alpha," or skill.
- A slope coefficient against the excess returns of peer group, often referred to as "beta."

The alpha term represents the manager's excess return against the peer group. The beta term measures the extent to which the manager employs greater or lesser amounts of leverage than do competitors.

Standard confidence tests can be applied to the regression's outputs. The alpha term can be tested for statistical significance to see if it is both positive and statistically different from zero.

This performance tool incorporates the following strengths:

- It allows management to opine whether skill is truly present or excess returns are happenstance. It tests whether the manager has generated excess returns vis à vis the peer group.
- It allows management to distinguish between excess returns due to leverage and excess returns due to skill.
- The alpha and beta statistics, and tests of significance, are easy to calculate.

This performance tool incorporates the following weaknesses:

- There may not be a sufficient number of data points to permit a satisfactory conclusion about the statistical significance of alpha or beta.
- Returns of the peer group are biased due to the existence of survivorship biases.
- There is often a wide divergence in the amount of money under management among the peers. It is often easier to make larger risk-adjusted excess returns with smaller sums under management than with larger sums.

SUMMARY

Risk represents a shadow cost that businesses accept in order to produce profit. For a return to be deemed acceptable, expected returns must be adequate to compensate for the risk assumed. Risk management therefore implies that cost benefit process is at work.

Risk is a scarce resource in the sense that organizations place limits on their willingness to accept loss. For any given level of risk assumed, the objective is to engage into as many intelligent profit-making opportunities as possible. If risk is squandered or used unwisely, the ability of the organization to achieve its profit objectives is put at risk. If excessive levels of risk are taken vis à vis budget, the organization is risking unacceptably large losses in order to produce returns that it neither expects nor desires. If too little risk is taken vis à vis budgeted levels, return expectations will likely fall short of budget. The ability of an organization to achieve its risk and return targets is put at risk anytime that risk capital is used wastefully or in amounts inconsistent with the policies established by such organization.

There are three fundamental dimensions behind risk management—planning, budgeting, and monitoring. We observe that these three dimensions are intimately related and that they can be more completely understood by looking at their commonly used counterparts in the world of financial accounting controls. We posit that there is a direct correspondence between financial planning, financial budgeting, and financial variance monitoring and their risk management counterparts—namely, risk planning, risk budgeting, and risk monitoring. This conclusion follows from the assertion that risk is the shadow cost behind returns. Hence behind every line item in a financial plan or budget must lie a corresponding risk dimension. Financial plans and budgets can therefore be alternatively expressed using risk management vocabulary.

The risk plan should set points of success or failure for the organization (e.g., return and volatility expectations, VaR policies, risk diversification standards, minimum acceptable levels of return on risk capital, etc.). The risk plan should be well vetted and discussed among the organization's senior leadership and oversight bodies. Its main themes should be capable of being articulated to analysts, boards, actuaries, management teams, and so on. For example, strategic plans have ROE targets and business diversification policies that are well known. The risk plan should describe how risk capital is to be allocated such that the expected returns on such risk capital yield the financial outcomes sought with a high degree of certainty.

The risk budget—often called asset allocation—quantifies the vision of the risk plan. The risk budget is a numeric blueprint that gives shape and form to the risk plan. There are many similarities between financial budgets and risk budgets. Financial budgets calculate net income as the difference between revenue and expenses. ROE is then estimated as net income divided by capital invested. In the case of risk budgets, a risk "charge"—defined as VaR or some other proxy for "risk expense"—can be associated with each line item of projected revenue and expense. Hence, a RORC (return on risk capital) can be associated with each activity as well as for the aggregation of all activities. In the case of both financial and risk budgets, ROE and RORC must exceed some minimum levels for them to be deemed acceptable. Both statistics are concerned with whether the organization is sufficiently compensated—in

cost/benefit terms—for the expenses and/or risks associated with generating revenues. Finally, both RORC and ROE can and should be estimated over all time intervals that are deemed relevant.

If we accept the premise that risk capital is a scarce commodity, it follows that monitoring controls should exist to ensure that risk capital is used in a manner consistent with the risk budget. Material variances from risk budget are threats to the investment vehicle's ability to meet its ROE and RORC targets. If excessive risk is used, unacceptable levels of loss may result. If too little risk is spent, unacceptable shortfalls in earnings may result. Risk monitoring is required to ensure that material deviations from risk budget are detected and addressed in a timely fashion. The chapter introduces the concept of an independent risk management unit (RMU) as a best practice in risk monitoring space. It discusses its objectives and provides examples of how it might operate in practice.

The final part of the chapter deals with performance measurement tools and related theory. Performance tools are especially robust when they confirm a priori expectations regarding the quality of returns. Among the objectives of these tools are:

- To determine whether a manager generates consistent excess risk-adjusted performance vis à vis a benchmark.
- To determine whether a manager generates superior risk-adjusted performance vis à vis the peer group.
- To determine whether the returns achieved are sufficient to compensate for the risk assumed in cost/benefit terms.
- To provide a basis for identifying those managers whose processes generate high-quality excess risk-adjusted returns. We believe that consistently superior risk-adjusted performance results suggest that a manager's processes, and the resulting performance, can be replicated in the future, making the returns high-quality.

The chapter then describes tools to measure the nature of performance. Unusual volatility and performance results can be identified by categorizing each outcome as statistically expected (a green zone outcome), somewhat unusual (a yellow zone outcome), and statistically improbable (a red zone outcome). Other performance tools that are explored include return attribution, the Sharpe and information ratios, and portfolio manager alpha versus the benchmark and versus a peer group. In each case, strengths and weaknesses of the performance measurement tool are briefly discussed.

Appendix B provides a more mathematical treatment of account performance measurement.

APPENDIX A

Representative Questions to Help Define Manager Philosophies/Processes

1. What sectors do you trade?
2. What countries and regions do you trade?
3. What products do you trade (equities, over-the-counter (OTC) foreign exchange (FX), fixed income, etc.)?
4. If you trade OTC, do ISDA, FX Netting agreements, and so on exist?
5. How many accounts do you trade?
6. Define your assets under management.
7. Are you able to produce a historical track record?
8. Does your strategy require a minimum amount of money under management in order for you to trade your entire portfolio?
9. Is your process capacity constrained? Can you estimate at what point it might be?
10. Describe the process by which you know that you are trading in accordance with client guidelines.
11. Do you believe that your process is volume sensitive in terms of the number of accounts under management? If so, discuss.
12. Describe how your process generates profits. That is, what is the source of your excess returns (e.g., superior stock selection, superior quantitative modeling, superior fundamental research, etc.)?
13. Define the list of your benchmarks. Are all of them easily calculated or are some nonstandard? For nonstandard benchmarks, describe how you manage risk in your portfolio. Would you prefer standard benchmarks if that option was available to you?
14. What risk system do you use to measure risk and build portfolios?
15. Have you found weaknesses or problems with these systems from time to time? To the extent that these systems can be inadequate, how do you compensate?
16. Define the following on a daily, monthly, quarterly, and annual basis both in terms of active weights vis à vis a benchmark as well as in terms of marginal contribution to risk: maximum exposure by security; maximum exposure by

sector; maximum exposure by country; maximum exposure at the portfolio level.

A. For each of the above, define exposure at the one and three standard deviation levels.

B. When will you liquidate a position? Does this answer correlate to the answers given at (A) above?

C. At what point are losses vis à vis the benchmark so large that you would conclude that your process is no longer working?

17. Describe those environments that are harmful for you.

18. Describe those environments that are favorable for you.

19. Is any part of your book vulnerable to market illiquidity? That is, does the genre of products you trade have evidence of becoming much less illiquid (based on historical observation)?

20. Do you have risk limits in terms of:

- Maximum percentage of the security outstanding
- Maximum percentage of daily volume (alternatively, how many days to liquidate if you never want to be more than, say, 15 percent of the daily volume).

Describe how these limits are applied. Are they applied on an account-by-account basis as well as on an overall basis (i.e., the sum total of all accounts under your direction)?

21. Define the risk factors that drive your returns. Does your risk software follow all of these factors? If not, how do you compensate?

22. Describe the process by which you review your daily results. What reports do you look at?

23. What process exists to ensure that accounts are traded in a parallel fashion?

24. Of the various fundamental factors followed by your risk system, define a normal band around each one.

25. Does redemption risk enter into your portfolio management? If so, how?

26. Have you had any material trading errors over the past year? If so, what were the circumstances?

27. At year-end, how would you define successful portfolio management? What statistics should we look to as guidance for measuring the quality of risk-adjusted performance?

28. Describe controls over valuation of your portfolio.

29. Describe the nature of the credit review you perform for custodians and executing brokers.

APPENDIX B

Calculation of Account Performance

Performance measurement provides an objective, quantitative assessment of the change in value of a portfolio or portfolio segment over an evaluation period, including the impact of any cash flows during that period. The calculation of total return in the absence of cash flows for a period is based on the formula

$$r_p(t)\frac{MVE - MVB}{MVB} \tag{B.1}$$

where $r_p(t)$ = Portfolio return

MVE = Market value of portfolio at end of period, including all accrued income

MVB = Portfolio's market value at beginning of period, including all income accrued up to end of previous period

This definition of a portfolio's return is valid only if there are no intraperiod cash flows. In practice, this condition is often violated as cash flows frequently occur due to capital allocated to or removed from the portfolio (client's account) or through transactions from buying and selling securities.

If cash flows do occur over the period in which returns are calculated, we need to do the following:

- Compute the market value of the cash flows at the date/time at which they occur.
- Calculate the interim rate of return for the subperiod according to Equation (B.1).
- Link the sub-period returns to get the return for the entire period.

In equity markets, the primary drivers of performance include the shares held of each asset and its market price as well as accrued income from dividends. Dividends ex-not-paid affect a stock's price whereas cash dividends on the pay date do not.

When cash flows occur, there are two proposed methods for measuring a portfolio's return. The first is a *dollar-weighted return* and the second is a *time-weighted return*.

Dollar-Weighted Return

There are two methods for computing a dollar-weighted return. The first is the internal rate of return and the second is the modified Dietz method. To compute the internal rate of return of a portfolio we assume that the portfolio has $I(I = 1, \ldots, I)$ cash flows over some period (e.g., one day, one month, one quarter)

and solve for the internal rate of return, IRRATE, such that the following relationship holds

$$MVE = \sum_{i=1}^{I} FLOW_i \times (1 + IRRATE_i)^{\overline{w}_i} \qquad \text{(B.2)}$$

where $FLOW_i$ = $i$thcash flow over the return period, in the form of either a deposit (cash or security) or a withdrawal

$\overline{w}_i$ = Proportion of the total number of days in period that $FLOW_i$ has been in (or out of) portfolio. The formula for $\overline{w}_i$ assuming cash flows occur at end of day, is

$$\overline{w}_i = \frac{(CD - D_i)}{CD}$$

where CD = Total number of days in return period

D_i = Number of days since beginning of period when the flow, $FLOW_i$, occurred

Equation (B.2) is also known as the modified Bank Administration Institute method (modified BAI). It is an acceptable approximation to the time-weighted return (discussed in the next section) when the results are calculated at least quarterly and geometrically linked over time.

A portfolio's return based on the Modified Dietz method is given by

$$R_{\text{Dietz}} = \frac{MVE - MVB - F}{MVB + FW} \qquad \text{(B.3)}$$

where F = Sum of cash flows within period

FW = Sum of cash flows each multiplied by its weight

$$\left(\text{i.e., } FW = \sum_{i=1}^{I} FLOW_i \times \overline{w}_i\right)$$

Time-Weighted Return

Ideally, we would want to compute a portfolio's return in such a way as to incorporate the precise time when the cash flows occur. To this end, the time-weighted rate of return (also known as the daily valuation method) for a portfolio is given by

$$R_{RWR} = (S_1 \times S_2 \times \ldots \times S_P) - 1 \qquad \text{(B.4)}$$

where $P(p = 1, \ldots, P)$ is the number of subperiods that are defined within the period's return and

$$S_p = \frac{MVE_p}{MVB_p} \qquad \text{(B.5)}$$

where MVE_p is the market value of the portfolio at the end of the pth subperiod, before any cash flows in period p but including accrued income for the period, and MVB_p is the market value at the end of the previous subperiod (i.e., beginning of this subperiod), including any cash flows at the end of the previous subperiod and including accrued income up to the end of the previous period. This method is the most exact of the three explained here.

Note that the main difference between the dollar-weighted return and the time-weighted return is that the former assumes the same rate of return over the whole period. The time-weighted return, on the other hand, uses the geometric average of returns from each individual period.

A good way to understand the methods described is to look at a numerical example. Suppose that on January 1, 2002, we invested \$100 in the Nasdaq Composite index. On March 1, 2002, we invest another \$100. The total return on the Nasdaq from January 1, 2002, through February 28, 2002, was −11.22 percent. Hence our initial investment of \$100 is now worth \$88.78. However, since we invested another \$100, the total value of our investment is \$188.78. By March 28, 2002, the total value of our investment has grown to \$201.20 and we sell \$100. The Nasdaq then declines until finally, on May 10, 2002, we are left with \$87.79.

We compute our return on this investment as of May 10, 2002, under the different methods presented above.

- The ideal time-weighted return is

$$[(88.78/100) \times (201.20/188.78) \times (87.79/101.20)] - 1 = -17.92\%$$

- The dollar-weighted annualized return based on the BAI method is

$$\begin{aligned} 87.79 &= 100(1 + IRRATE)^{90/252} \\ &\quad + 100(1 + IRRATE)^{50/252} \\ &\quad - 100(1 + IRRATE)^{30/252} \\ IRRATE &= -25.50\% \end{aligned}$$

- According to the modified Dietz method, the annualized return is

$$\frac{87.79 - 100 - (100 - 100)}{100 + 7.94} = -11.31\%$$

Clearly, the dollar-weighted return calculation takes into account the timing of the decisions to sell or buy as reflected by the −25.50 percent return.

Computing Returns

Let $R_n^\ell(t)$ represent the local return on the nth asset as measured in percent:

$$R_n^\ell(t) = \frac{P_n^\ell(t) + d_n(t-h,t) - P_n^\ell(t-1)}{P_n^\ell(t-1)} \tag{B.6}$$

$P_n^\ell(t) =$ Time t local price of security or asset

$d_n(t-h,t) =$ Dividend (per share) paid out at time t for period $t-h$ through t

In a global framework we need to incorporate exchange rates into the return calculations. We define exchange rates as the reporting currency over the local currency (reporting/local). The local currency is sometimes referred to as the risk currency. For example, USD/GBP would be the exchange rate where the reporting currency is the U.S. dollar and the risk currency is the British pound. A USD-based investor with holdings in U.K. equities would use the USD/GBP rate to convert the value of the stock to U.S. dollars.

Suppose a portfolio with U.S. dollars as its reporting currency has holdings in German, Australian, and Japanese equities. The local and/or risk currencies are EUR, AUD, and JPY, respectively. The total return of each equity position consists of the local return on equity and the return on the currency expressed in reporting/local.

We assume that a generic portfolio contains N assets ($n = 1, \ldots, N$). Let $P_n^\ell(t)$ represent the price, in euros, of one share of Siemens stock. $X_{ij}(t)$ is the exchange rate expressed as the ith currency per unit of currency j. For example, with USD as the reporting currency, the exchange rate where $X_{ij}(t) = \text{USD}/\text{EUR}$ (i is USD and j is EUR) is used to convert Siemens equity (expressed in euros) to U.S. dollars. In general, the exchange rate is expressed in reporting over local currency.

It follows from these definitions that the price of the nth asset expressed in reporting currency is

$$P_n(t) = P_n^\ell(t) \times X_{ij}(t) \tag{B.7}$$

We use (B.7) as a basis for defining total return, local return, and exchange rate return. The total return of an asset or portfolio is simply the return that incorporates both the local return and exchange rate return. Depending on how returns are defined—continuous or discrete (percent)—we get different equations for how returns are calculated. Following directly from (B.7), an asset's total return, using percent returns, is defined as

$$\begin{aligned} R_n(t) &= [1 + R_n^\ell(t)][1 + E_{ij}(t)] - 1 \\ &= R_n^\ell(t) + E_{ij}(t) + R_n^\ell(t) \times E_{ij}(t) \end{aligned} \tag{B.8}$$

where $R_n(t) =$ One-period total return on the nth asset

$R_n^\ell =$ One-period percent return on the equity positions expressed in local currency (i.e., the local return)

$E_{ij}(t) =$ One-period percent return on the ith currency per unit of currency j

$$E_{ij}(t) = X_{ij}(t)/X_{ij}(t-1) - 1$$

For example, suppose that the nth position is a position in the DAX equity index. In this case, $R_n^\ell(t)$ is the local return on DAX and $E_{ij}(t)$ is the return in the USD/EUR exchange rate. When the euro strengthens, USD/EUR increases and $E_{ij}(t) > 0$. Holding all other things constant, this increases the total return on the equity position.

8

Portfolio Performance Evaluation

Learning Objectives

After completing this reading you should be able to:

- Differentiate between the time-weighted and dollar-weighted returns of a portfolio and describe their appropriate uses.
- Describe risk-adjusted performance measures, such as Sharpe's measure, Treynor's measure, Jensen's measure (Jensen's alpha), and the information ratio, and identify the circumstances under which the use of each measure is most relevant.
- Describe the uses for the Modigliani-squared and Treynor's measure in comparing two portfolios and the graphical representation of these measures.
- Determine the statistical significance of a performance measure using standard error and the t-statistic.
- Describe style analysis.
- Explain the difficulties in measuring the performance of actively managed portfolios.
- Describe performance manipulation and the problems associated with using conventional performance measures.
- Describe techniques to measure the market timing ability of fund managers with a regression and with a call option model and compute return due to market timing.
- Describe and apply performance attribution procedures, including the asset allocation decision, sector and security selection decision, and the aggregate contribution.

Excerpt is Chapter 24 of Investments, *Twelfth Edition, by Zvi Bodie, Alex Kane, and Alan J. Marcus.*

MOST FINANCIAL ASSETS are managed by professional investors, who thus at least indirectly allocate the lion's share of capital across firms. Efficient allocation therefore depends on the quality of these professionals and the ability of financial markets to identify and direct capital to the best stewards. Therefore, if capital markets are to be reasonably efficient, investors must be able to measure the performance of their asset managers.

How can we evaluate investment performance? It turns out that even an average portfolio return is not as straightforward to measure as it might seem. In addition, adjusting average returns for risk presents a host of other problems. In the end, performance evaluation is far from trivial.

We begin this chapter with the measurement of portfolio returns. From there we move on to conventional approaches to risk adjustment. We consider the situations in which each of the standard risk-adjusted performance measures will be of most interest to investors and show how style analysis may be viewed as a generalization of the index model and the alpha statistic that it generates.

Performance measurement becomes far more difficult when managers change portfolio composition during the measurement period, so we also examine the complications posed by changes in the risk characteristics of the portfolio. One particular way in which this may occur is when managers attempt to time the broad market and adjust portfolio beta in anticipation of market movements. Market timing raises a wide range of issues in performance evaluation.

We close the chapter with a look at performance attribution techniques. These are tools used to decompose managers' performance into results that can be attributed to security selection, sector selection, and asset allocation decisions.

8.1 THE CONVENTIONAL THEORY OF PERFORMANCE EVALUATION

Average Rates of Return

Suppose we evaluate the performance of a portfolio over a period of 20 years. The *arithmetic average* return is the sum of the 20 annual returns divided by 20. In contrast, the *geometric average* is the constant annual return over the 20 years that would provide the same total cumulative return over the entire investment period. Therefore, the *geometric average*, r_G, is defined by

$$(1 + r_G)^{20} = (1 + r_1)(1 + r_2) \cdots (1 + r_{20})$$

The right-hand side of this equation is the compounded final value of a $1 investment earning the 20 annual rates of return. The left-hand side is the compounded value of a $1 investment earning r_G *each year*. We solve for $1 + r_G$ as

$$1 + r_G = [(1 + r_1)(1 + r_2) \cdots (1 + r_{20})]^{1/20}$$

Each return has an equal weight in the geometric average. For this reason, the geometric average is referred to as a **time-weighted average.**

Time-Weighted Returns versus Dollar-Weighted Returns

To set the stage for the more subtle issues that follow, let's start with a trivial example. Consider a stock paying a dividend of $2 annually that currently sells for $50. You purchase the stock today, collect the $2 dividend, and then sell the stock for $53 at year-end. Your rate of return is

$$\frac{\text{Total proceeds}}{\text{Initial investment}} = \frac{\text{Income} + \text{Capital gain}}{50} = \frac{2 + 3}{50} = .10, \text{ or } 10\%$$

Another way to derive the rate of return that is useful in the more difficult multiperiod case is to set up the investment as a discounted cash flow problem. Call r the rate of return that equates the present value of all cash flows from the investment with the initial outlay. In our example, the stock is purchased for $50 and generates cash flows at year-end of $2 (dividend) plus $53 (sale of stock). Therefore, we solve $50 = (2 + 53)/(1 + r)$ to find again that $r = .10$, or 10%.

When we consider investments over a period during which cash was added to or withdrawn from the portfolio, measuring the rate of return becomes more difficult. To continue our example, suppose that you purchase a second share of the same stock at the end of the first year, and hold both shares until the end of year 2, at which point you sell each share for $54.

Total cash outlays are shown below

Time	Outlay
0	$50 to purchase first share
1	$53 to purchase second share a year later
	Proceeds
1	$2 dividend from initially purchased share
2	$4 dividend from the 2 shares held in the second year, plus
	$108 received from selling both shares at $54 each

Using the discounted cash flow (DCF) approach, we can solve for the average return over the two years by equating the present values of the cash inflows and outflows:

$$50 + \frac{53}{1 + r} = \frac{2}{1 + r} + \frac{112}{(1 + r)^2}$$

resulting in $r = 7.117\%$. This is the internal rate of return on the investment.[1]

The internal rate of return is called the **dollar-weighted rate of return**. It is "dollar weighted" because the stock's performance in the second year, when two shares of stock are held, has a greater influence on the average overall return than the first-year return, when only one share is held.

The time-weighted (geometric average) return is 7.81%:

$$r_1 = \frac{53 + 2 - 50}{50} = .10 = 10\% \quad r_2 = \frac{54 + 2 - 53}{53} = 0.566 = 5.66\%$$

$$r_G = (1.10 \times 1.0566)^{1/2} - 1 = .0781 = 7.81\%$$

The dollar-weighted average is less than the time-weighted average in this example because the return in the second year, when more money was invested, is lower.

Concept Check 8.1 Shares of XYZ Corp, pay a $2 dividend at the end of every year on December 31. An investor buys two shares of the stock on January 1 at a price of $20 each, sells one of those shares for $22 a year later on the next January 1, and sells the second share an additional year later for $19. Find the dollar- and time-weighted rates of return on the 2-year investment.

Adjusting Returns for Risk

Evaluating performance based on average return alone is not very useful because returns must be adjusted for risk before they can be compared meaningfully. The simplest and most popular way to make the adjustment is to compare rates of return with those of other investment funds with similar risk characteristics. For example, high-yield bond portfolios are grouped into one **comparison universe,** growth stock equity funds are grouped into another, and so on. Then the (usually time-weighted) average returns of each fund within the universe are ordered, and each portfolio manager receives a percentile ranking of relative performance within the comparison group. For example, the manager with the ninth-best performance in a universe of 100 funds would be the 90th percentile manager: Her performance was better than 90% of all competing funds over the evaluation period.

These relative rankings are usually displayed in a chart such as that in Figure 8.1. The chart summarizes performance rankings over four periods: 1 quarter, 1 year, 3 years, and 5 years. The top and bottom lines of each box are drawn at the rate of return of the 95th and 5th percentile managers. The three dashed lines correspond to the rates of return of the 75th, 50th (median), and 25th percentile managers. The diamond is drawn at the average return of a particular fund and the square is drawn at the return of a benchmark index, such as the S&P 500. The placement of the diamond within the box is an easy-to-read representation of the performance of the fund relative to the comparison universe.

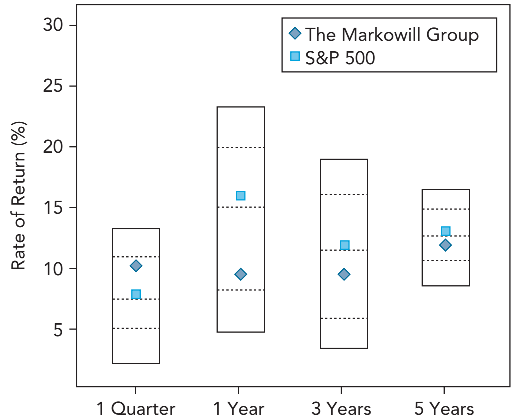


Figure 8.1 **Universe comparison, periods ending December 31, 2025.**

This comparison of performance with other managers of similar investment style is a useful first step in evaluating performance. However, such rankings can be misleading. Within a particular universe, some managers may concentrate on particular subgroups, so that portfolio characteristics are not truly comparable. For example, within the equity universe, one manager may concentrate on high-beta or aggressive growth stocks. Similarly, within fixed-income universes, durations can vary across managers.

A more fundamental problem with the comparison universe methodology is that the benchmark performance is not an investable strategy. When we evaluate the performance of investment professionals and implicitly ask if they have justified their expenses, we would like to compare their performance to what would have been available to a passive investor. Market index portfolios, many of which are available as mutual funds or ETFs, are natural benchmarks because investors can use these products to match the return on the index. In contrast, the median fund in Figure 8.1 cannot be known in advance. Because passive investors cannot reliably lock in the median return, it is an unsatisfactory benchmark.

These considerations suggest that a more precise means for risk adjustment is desirable.

[1] Excel's function XIRR calculates IRR. The function provides the IRR between any two dates given a starting value, cash flows at various dates in between (with additions given as negative numbers, and withdrawals as positive values), and a final value on the closing date.

Risk-Adjusted Performance Measures

Methods of risk-adjusted performance evaluation using mean-variance criteria came on stage simultaneously with the capital asset pricing model. Jack Treynor,[2] William Sharpe,[3] and Michael Jensen[4] recognized immediately the implications of the CAPM for rating the performance of managers. Within a short time, academicians were in command of a battery of performance measures, and a bounty of scholarly investigation of mutual fund performance was pouring from ivory towers. Shortly thereafter, agents emerged who were willing to supply rating services to portfolio managers and their clients.

But while widely used, risk-adjusted performance measures each have their own limitations. Moreover, their reliability requires quite a long history of consistent management with a steady level of performance and a representative sample of investment environments: bull as well as bear markets.

We start by cataloging some common risk-adjusted performance measures for a portfolio, *P*, and examine the circumstances in which each might be most relevant.

1. *Sharpe ratio:* $(\bar{r}_P - \bar{r}_f)/\sigma_P$

   The Sharpe ratio divides average portfolio excess return over the sample period by the standard deviation of returns over that period. It measures the reward to (total) volatility trade-off.[5]

2. *Treynor measure:* $(\bar{r}_P - \bar{r}_f)/\beta_P$

   Like the Sharpe ratio, **Treynor's measure** gives excess return per unit of risk, but it uses systematic risk instead of total risk.

3. *Jensen's alpha:* $\alpha_P = \bar{r}_P - [\bar{r}_f + \beta_P(\bar{r}_M - \bar{r}_f)]$

   Jensen's alpha is the average return on the portfolio over and above that predicted by the CAPM, given the portfolio's beta and the average market return.[6]

4. *Information ratio:* $\alpha_P/\sigma(e_P)$

   The **information ratio** divides the alpha of the portfolio by nonsystematic risk, called "tracking error" in the industry. It measures abnormal return per unit of risk that in principle could be diversified away by holding a market index portfolio. (We should note that industry jargon tends to be a little loose concerning this topic. Some define the information ratio as excess return—rather than alpha—per unit of nonsystematic risk, using *appraisal ratio* to refer to the ratio of alpha to nonsystematic risk. Unfortunately, terminology in the profession is not fully uniform, and you may well encounter both of these definitions of the information ratio. We will consistently define it as we have done here, specifically as the ratio of alpha to the standard deviation of residual returns.)

Each performance measure has some appeal. But as Concept Check 8.2 shows, these competing measures do not necessarily provide consistent assessments of performance because the risk measures used to adjust returns differ substantially. Therefore, we need to consider the circumstances in which each of these measures is appropriate.

Concept Check 8.2 Consider the following data for a particular sample period:

	Portfolio *P*	Market *M*
Average return	35%	28%
Beta	1.20	1.00
Standard deviation	42%	30%
Tracking error (nonsystematic risk), $\sigma(e)$	18%	0

Calculate the following performance measures for portfolio *P* and the market: Sharpe, Jensen (alpha), Treynor, information ratio. The T-bill rate during the period was 6%. By which measures did portfolio *P* outperform the market?

The Sharpe Ratio for Overall Portfolios

If investor preferences can be summarized by a mean-variance utility function, we can arrive at a relatively simple criterion. The particular utility function that we used is

$$U = E(r_P) - ½ A\sigma_P^2$$

where *A* is the coefficient of risk aversion. With mean-variance preferences, the investor wants to maximize the Sharpe ratio $[E(r_P) - r_f]/\sigma_P$. The problem reduces to the search for the portfolio with the highest possible Sharpe ratio. The Sharpe ratio is the slope of the capital allocation line, and investors will naturally seek the portfolio that provides the greatest slope, that is,

[2] Jack L. Treynor, "How to Rate Management Investment Funds," *Harvard Business Review* 43 (January–February 1966).

[3] William F. Sharpe, "Mutual Fund Performance," *Journal of Business* 39 (January 1966).

[4] Michael C. Jensen, "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance*, May 1968; and Michael C. Jensen, "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business*, April 1969.

[5] We place bars over r_f as well as r_p to denote the fact that because the risk-free rate may not be constant over the measurement period, we are taking a sample average, just as we do for r_p. Equivalently, we may simply compute sample average *excess* returns.

[6] In many cases performance evaluation assumes a multifactor market. For example, when the Fama-French 3-factor model is used, Jensen's alpha will be $\alpha_P = \bar{r}_p - \bar{r}_f - \beta_P(\bar{r}_M - \bar{r}_f) - s_P\bar{r}_{SMB} - h_p\bar{r}_{HML}$, where s_P is the loading on the SMB portfolio and h_P is the loading on the HML portfolio.

the greatest increase in expected return for each unit increase in volatility.

We focus here on total volatility rather than systematic risk because we are looking at the full portfolio rather than a small component of it. The benchmark for acceptable performance is the Sharpe ratio of the market index because the investor can easily opt for a passive strategy by investing in an indexed equity mutual fund. The actively managed portfolio therefore must offer a higher Sharpe ratio than the market index if it is to be an acceptable candidate for the investor's optimal risky portfolio.

The M^2 Measure and the Sharpe Ratio

While the Sharpe ratio can be used to rank portfolio performance, its numerical value is not easy to interpret. Comparing the ratios for portfolios *M* and *P* in Concept Check 8.2, you should have found that $S_P = .69$ and $S_M = .73$. This suggests that portfolio P underperformed the market index. But is a difference of .04 in the Sharpe ratio economically meaningful? We often compare rates of return, but these values are difficult to interpret.

An equivalent representation of Sharpe's ratio was proposed by Graham and Harvey, and later popularized by Leah Modigliani of Morgan Stanley and her grandfather Franco Modigliani, past winner of the Nobel Prize in Economics.[7] Their approach has been dubbed the M^2 measure (for Modigliani-squared). Like the Sharpe ratio, M^2 focuses on total volatility as a measure of risk, but its risk adjustment leads to an easy-to-interpret differential return relative to the benchmark index.

To compute M^2, we imagine that an active portfolio, *P*, is mixed with a position in T-bills so that the resulting "adjusted" portfolio matches the volatility of a passive market index such as the S&P 500. For example, if the active portfolio has 1.5 times the standard deviation of the index, you would mix it with T-bills using proportions of 1/3 in bills and 2/3 in the active portfolio. The resulting adjusted portfolio, which we call *P**, would then have the same standard deviation as the index. (If the active portfolio had *lower* standard deviation than the index, it would be leveraged by borrowing money and investing the proceeds in the portfolio.) Because the market index and portfolio *P** have the same standard deviation, we may compare their performance simply by comparing returns. This is the M^2 measure for portfolio *P:*

$$M_P^2 = r_{P^*} - r_M \tag{8.1}$$

[7] John R. Graham and Campbell R. Harvey, "Grading the Performance of Market Timing Newsletters," *Financial Analysts Journal* 53 (November/December 1997), pp. 54–66; and Franco Modigliani and Leah Modigliani, "Risk-Adjusted Performance," *Journal of Portfolio Management*, Winter 1997, pp. 45–54.

Example 8.1 M^2 Measure

Consider the performance of portfolio *P* and the market index portfolio, *M* from Concept Check 8.2. Portfolio *P* had the higher return, but it also had higher risk, with the result that its Sharpe ratio was less than the market's. In Figure 8.2, we plot the average return and volatility of each portfolio and draw the capital allocation line for each. Portfolio *P*'s CAL is less steep than the CML, consistent with its lower Sharpe ratio.

The adjusted portfolio *P** is formed by mixing bills with portfolio *P*. We use weights $30/42 = .714$ in *P* and $1 - .714 = .286$ in T-bills. By construction, the adjusted portfolio has exactly the same standard deviation as the market index: $30/42 \times 42\% = 30\%$. Despite its equal volatility, its return would be only $(.286 \times 6\%) + (.714 \times 35\%) = 26.7\%$, which is less than that of the market index.

You can see in Figure 8.2 that portfolio *P** is formed by moving down portfolio *P*'s capital allocation line (by mixing *P* with T-bills) until we reduce the standard deviation by just enough to match that of the market index. M^2 is the vertical distance between points *M* and *P**. The M^2 of of portfolio *P* is therefore $26.7\% - 28\% = -1.3\%$; the negative M^2 is consistent with the inferior Sharpe ratio.

The Treynor Ratio

In many circumstances, you may have to select funds or portfolios that will be mixed together to form the investor's overall risky portfolio. For example, the manager in charge of a large pension plan might parcel out the total assets to several portfolio managers. Consider CalPERS (the California Public Employee

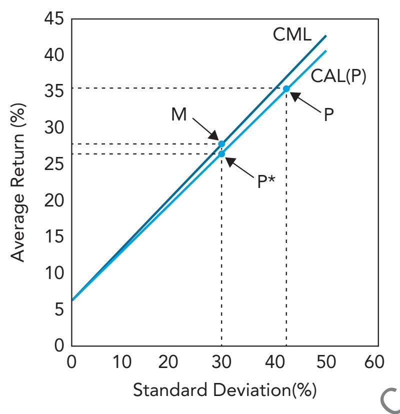


Figure 8.2 **M^2 of portfolio *P* is negative even though its average return was greater than that of the market index, *M*.**

Retirement System), which had a portfolio of about $350 billion in early 2019. Like many large plans, it uses a *funds of funds* approach, allocating the endowment among a number of professional managers (funds). How should you compare performance across candidate managers in this context?

When employing a number of managers, the nonsystematic risk of each manager will be largely diversified away, so only systematic risk is relevant. The appropriate performance metric when evaluating *components* of the full risky portfolio is now the Treynor measure: This reward-to-risk ratio divides expected excess return by *systematic* risk (i.e., by beta).

Suppose the relevant data for two portfolios and the market index are given in Table 8.1. Portfolio Q has an alpha of 3.5%, while that of portfolio *U* is 3%. It might appear that you would prefer portfolio Q but this turns out not to be the case. To see why, turn to Figure 8.3. As in Figure 8.2, we plot portfolios on a return-risk graph, but now measure risk using beta.

Table 8.1 Portfolio Performance

	Risk-free Asset	Portfolio Q	Portfolio *U*	Market Index, *M*
Beta	0	1.3	0.8	1.0
Average return	6	22.0	17.0	16.0
Excess return (%)	0	16.0	11.0	10.0
Alpha (%)	0	3.5	3.0	0.0

Notes: Excess return = Average return − Risk-free rate
Alpha = Average return − Beta × (Market return − Risk-free rate)

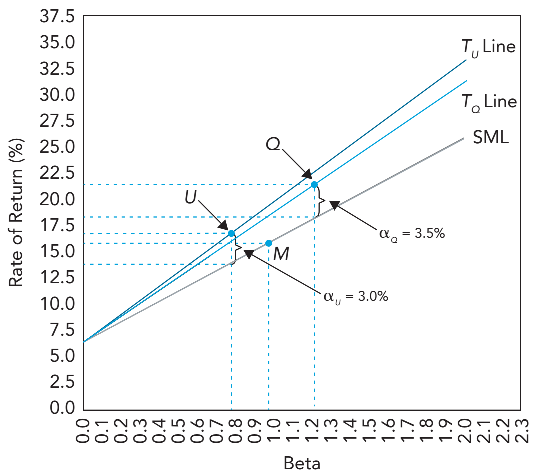


Figure 8.3 Treynor measures of two portfolios and the market index.

The SML is the black line from the risk-free rate on the vertical axis through point *M*. Its slope is the market's Treynor ratio, the increase in return[8] per increase in beta offered by portfolios that mix the market index with T-bills.

The line through point Q shows us all the beta-return combinations that can be achieved by mixing bills with portfolio Q. The slope of this line is portfolio Q's Treynor ratio, so we call it the T_Q-line. Similarly, the T_U-line goes from the origin through point *U*.

As always, the investor wants the portfolio that provides the best risk-return trade-off. When we measure risk using beta, that will be the portfolio with the steepest *T*-line, or, equivalently, the portfolio with the highest Treynor ratio. Therefore, we see that despite its lower alpha, portfolio *U* actually will be preferred to portfolio Q. For any level of risk (beta), it provides the higher return.

We can measure the advantage of each portfolio compared to the market index using a measure similar to M^2. We will again combine portfolios with T-bills to match market risk so that we can directly compare rates of return. Here, however, we use beta to measure risk, so we will form adjusted portfolios designed to match the beta of the market index (which, of course, is 1). We form the adjusted portfolio Q* by mixing Q with T-bills in proportions w_Q and $(1 - w_Q)$. We seek portfolio proportions that will make the beta of Q* equal to 1. Therefore, we choose W_Q to satisfy

$$\beta_{Q^*} = w_Q \times \beta_Q + (1 - w_Q) \times \beta_{\text{bills}} = w_Q \times 1.25 + (1 - w_Q) \times 0 = 1.0$$

which implies that $w_Q = .8$.

Because the systematic risk of the adjusted portfolio is constructed to match the market's, we can meaningfully compare their returns. The return on Q* is $.8 \times 22 + .2 \times 6 = 18.8\%$, which beats the market return by 2.8%. This measure is analogous to the M^2 measure, but it extends the Treynor rather than the Sharpe measure. Therefore, we will call it the *T*-square or T^2 measure. The following example computes the T^2 measure for portfolio *U*.

Example 8.2 Equalizing Beta and the T^2 Measure

Portfolio *U* has a beta of .8, which is less than the market beta of 1. Therefore, we will need to use leverage to increase systematic risk to match that of the market index. We choose w_U to satisfy:

$$\beta_{U^*} = w_U \times \beta_U + (1 - w_U) \times \beta_{\text{bills}} = w_U \times .8 + (1 - w_U) \times 0 = 1.0$$

[8] We talk about return rather than *expected* return here because we are envisioning ourselves comparing the actual performance of a portfolio manager who earned a given return and took on a given level of risk to the risk–return combinations that we could have earned by mixing the market index with a risk-free asset.

which implies that $w_U = 1.25$. So the adjusted portfolio (U^* *entails borrowing* at the risk-free rate, with all proceeds invested in portfolio U.

The return on the adjusted portfolio U^* is $1.25 \times 17 - .25 \times 6 = 19.75\%$, which beats the market's return by 3.75%. Portfolio U's higher T^2 measure compared to portfolio Q is consistent with the steeper slope it displayed in Figure 8.3. We can use either the T^2 or the Treynor measure to rank portfolios when beta is the relevant measure of risk.

The Information Ratio

Here is another situation that calls for yet another performance criterion. Consider a pension fund with a largely passive and well-diversified position—for example, a portfolio that resembles an indexed equity fund. Now the fund decides to add a position in an active portfolio to its current position. For example, a university might employ a hedge fund as a possible addition to its core positions in more traditional portfolios that were established primarily with concerns of diversification in mind.

When the hedge fund (or another active position) is optimally combined with the baseline indexed portfolio, the improvement in the Sharpe measure will be determined by its information ratio $\sigma_H/\sigma(e_H)$, according to

$$S_P^2 = S_M^2 + \left[\frac{\alpha_H}{\sigma(e_H)}\right]^2 \tag{8.2}$$

Equation 8.2 tells us that the appropriate performance measure for the hedge fund, H, is its information ratio (IR). If you are looking for active managers to add to a currently indexed position, you will want to select potential candidates with the best information ratios.

The information ratio is yet another version of a reward-to-risk ratio. In this context, the reward is the alpha of the active position. It is the expected return on that incremental portfolio over and above the risk premium that would normally correspond to its systematic risk. On the other hand, the incremental position tilts the total risky portfolio away from the passive index, and therefore exposes it to risk that could, in principle, be diversified. The information ratio quantifies the trade-off between alpha and diversifiable risk.

We can summarize the preceding discussion with the following table, which shows the definition of the various performance measures and the situations in which each is the relevant performance measure.

Performance Measure	Definition	Application
Sharpe	$\dfrac{\text{Excess return}}{\text{Standard deviation}}$	When choosing among portfolios competing for the overall risky portfolio
Treynor	$\dfrac{\text{Excess return}}{\text{Beta}}$	When ranking many portfolios that will be mixed to form the overall risky portfolio
Information ratio	$\dfrac{\text{Alpha}}{\text{Residual standard deviation}}$	When evaluating a portfolio to be mixed with the benchmark portfolio

The Role of Alpha in Performance Measures

Given the central role of alpha in the index model, the CAPM, and other models of risk versus return, you may be surprised that we have not encountered a situation in which alpha is the criterion used to choose one fund over another. But don't conclude from this that alpha does not matter! With some algebra, we can derive the relationship between the performance measures discussed so far and the alpha of the portfolio. The following table shows these relationships. In all cases, you can see that a positive alpha is necessary to outperform the passive market index. Because superior performance requires positive alpha, it is the most widely used performance measure.

	Treynor (T_p)	Sharpe* (S_p)	Information Ratio
Relation to alpha	$\dfrac{E(r_p) - r_f}{\beta_P} = \dfrac{\alpha_P}{\beta_P} + T_M$	$\dfrac{E(r_p) - r_f}{\sigma_P} = \dfrac{\alpha_P}{\sigma_P} + \rho S_M$	$\dfrac{\alpha_P}{\sigma(e_P)}$
Improvement compared to market index	$T_P - T_M = \dfrac{\alpha_P}{\beta_P}$	$S_P - S_M = \dfrac{\alpha_P}{\sigma_P} - (1-\rho)S_M$	$\dfrac{\alpha_P}{\sigma(e_P)}$

* ρ denotes the correlation coefficient between portfolio P and the market, and is less than 1.

However, while positive alpha is *necessary*, it is not sufficient to guarantee that a portfolio will outperform the index: Taking advantage of mispricing means departing from full diversification, which entails a cost in terms of nonsystematic risk. A mutual fund can achieve a positive alpha, yet, at the same time, its volatility may increase to a level at which its Sharpe ratio will actually fall.

EXCEL APPLICATIONS Performance Measurement

The following performance measurement spreadsheet computes all the performance measures discussed in this section. You can see how relative ranking differs according to the criterion selected. This Excel model is available in Connect or through your course instructor.

Excel Questions

1. Examine the performance measures of the funds included in the spreadsheet. Rank performance and determine whether the rankings are consistent using each measure. What explains these results?
2. Which fund would you choose if you were considering investing the entire risky portion of your portfolio? What if you were considering adding a small position in one of these funds to a portfolio currently invested in the market index?

	A	B	C	D	E	F	G	H	I	J	K
1	Performance Measurement							LEGEND			
2								Enter data			
3								Value calculated			
4								See comment			
5											
6					Non-						
7		Average	Standard	Beta	systematic	Sharpe's	Treynor's	Jensen's	M2	T2	Information
8	Fund	Return	Deviation	Coefficient	Risk	Measure	Measure	Measure	Measure	Measure	Ratio
9	Alpha	28.00%	27.00%	1.7000	5.00%	0.8148	0.1294	–0.0180	–0.0015	–0.0106	–0.3600
10	Omega	31.00%	26.00%	1.6200	6.00%	0.9615	0.1543	0.0232	0.0235	0.0143	0.3867
11	Omicron	22.00%	21.00%	0.8500	2.00%	0.7619	0.1882	0.0410	–0.0105	0.0482	2.0500
12	Millennium	40.00%	33.00%	2.5000	27.00%	1.0303	0.1360	–0.0100	0.0352	–0.0040	–0.0370
13	Big Value	15.00%	13.00%	0.9000	3.00%	0.6923	0.1000	–0.0360	–0.0223	–0.0400	–1.2000
14	Momentum Watcher	29.00%	24.00%	1.4000	16.00%	0.9583	0.1643	0.0340	0.0229	0.0243	0.2125
15	Big Potential	15.00%	11.00%	0.5500	1.50%	0.8182	0.1636	0.0130	–0.0009	0.0236	0.8667
16	S & P Index Return	20.00%	17.00%	1.0000	0.00%	0.8235	0.1400	0.0000	0.0000	0.0000	0.0000
17	T-Bill Return	6.00%		0.0000							
18											
19	Ranking By Sharpe's Measure				Non-						
20		Average	Standard	Beta	systematic	Sharpe's	Treynor's	Jensen's	M2	T2	Information
21	Fund	Return	Deviation	Coefficient	Risk	Measure	Measure	Measure	Measure	Measure	Ratio

Implementing Performance Measurement: An Example

To illustrate some of the calculations underlying portfolio evaluation, let's look at the performance of portfolio *P* over the last 12 months. Table 8.2 shows its excess return in each month as well as those of an alternative portfolio *Q* and the market index portfolio *M*. The bottom two rows in Table 8.2 provide the sample average and standard deviation of each portfolio. From these, and regressions of *P* and *Q* on *M*, we can compute the necessary performance statistics. These appear in Table 8.3.

Table 8.3 shows that portfolio *Q* is more aggressive than *P*, in the sense that its beta is significantly higher (1.40 versus .70). At the same time, *P*'s lower residual standard deviation indicates that it is better diversified than *Q* (2.02% versus 9.81%). Both portfolios outperformed the benchmark market index, as is evident from their higher Sharpe ratios (and thus positive M^2) and their positive alphas.

Which portfolio is more attractive based on reported performance? If *P* or *Q* represents the entire investment fund, *Q* would be preferable on the basis of its higher Sharpe measure (.49 vs. .43) and

Table 8.2 Excess Returns for Portfolios *P* and *Q* and the Market Index *M* Over 12 Months

Month	Portfolio *P*	Alternative *Q*	Index *M*
1	3.58%	2.81%	2.20%
2	−4.91	−1.15	−8.41
3	6.51	2.53	3.27
4	11.13	37.09	14.41
5	8.78	12.88	7.71
6	9.38	39.08	14.36
7	−3.66	−8.84	−6.15
8	5.56	0.83	2.74
9	−7.72	0.85	−15.27
10	7.76	12.09	6.49
11	−4.01	−5.68	−3.13
12	0.78	−1.77	1.41
Average	2.77	7.56	1.64
Standard deviation	6.45	15.55	8.84

00011693-00000002_CH08_p117-146.indd 124 05/07/22 3:07 PM

Table 8.3 Performance Statistics

	Portfolio *P*	Portfolio *Q*	Portfolio *M*
Sharpe ratio	0.43	0.49	0.19
M^2	2.16	2.66	0.00
SCL regression statistics			
Alpha	1.63	5.26	0.00
Beta	0.70	1.40	1.00
Treynor	3.97	5.38	1.64
T^2	2.34	3.74	0.00
$\sigma(e)$	2.02	9.81	0.00
Information ratio	0.81	0.54	0.00
R-square	0.91	0.64	1.00

better M^2 (2.66% vs. 2.16%). For the second scenario, where *P* and *Q* are competing for a role as one of a number of subportfolios, *Q* also dominates because its Treynor measure is higher (5.38 vs. 3.97). However, as an active portfolio to be mixed with the index portfolio, *P* is preferred because its information ratio [IR $= \alpha/\sigma(e)$] is higher (.81 vs. .54). Thus, the example illustrates that the appropriate way to evaluate a portfolio depends in large part on how the portfolio fits into the investor's overall investment plan.

This analysis is based on only 12 months of data, a period too short to lend statistical significance to the conclusions. Nevertheless, the analysis illustrates what one might wish to do with more extensive data. A model that calculates these performance measures is available in Connect.

Realized Returns versus Expected Returns

When evaluating a portfolio, the evaluator knows neither the portfolio manager's original expectations nor whether those expectations made sense. One can only observe performance after the fact and hope that the inherent "noise" in investment outcomes does not obscure true underlying ability. To avoid mistaken inferences, we have to determine the "significance level" of a performance measure to know whether it reliably indicates ability.

To illustrate, consider portfolio manager Joe Dart. Suppose that his portfolio has provided an alpha of 20 basis points per month, an impressive 2.4% per year before compounding. Let us assume that the return distribution of Joe's portfolio has constant mean, beta, and alpha, a heroic assumption, but one that is in line with the usual treatment of performance measurement. Suppose that for the measurement period, Joe's portfolio beta is 1.2 and the monthly standard deviation of the residual (nonsystematic risk) is $\sigma(e) = .02$ (i.e., 2% per month).

To estimate Joe's portfolio alpha from the security characteristic line (SCL), we regress portfolio excess returns on the market index. Suppose that we are in luck and (despite the underlying noise in investment returns) our regression estimates precisely match Joe's true parameters. That means that our SCL estimates for the *N* months are

$$\hat{\alpha} = .2\%, \quad \hat{\beta} = 1.2, \quad \hat{\sigma}(e) = 2\%$$

As outside evaluators who run such a regression, however, we do not know the true values. To assess whether the alpha estimate reflects true skill and not just luck due to statistical chance, we compute the *t*-statistic of the alpha estimate to determine whether we are justified in rejecting the hypothesis that Joe's true alpha is zero, that is, that he has no superior ability.

The standard error of the alpha estimate in the SCL regression is approximately

$$\hat{\sigma}(\alpha) = \frac{\hat{\sigma}(e)}{\sqrt{N}}$$

where *N* is the number of observations and $\hat{\sigma}(e)$ is the sample estimate of nonsystematic risk. The *t*-statistic for the estimate of alpha is then

$$t(\hat{\alpha}) = \frac{\hat{\alpha}}{\hat{\sigma}(\alpha)} = \frac{\hat{\alpha}\sqrt{N}}{\hat{\sigma}(e)} \tag{8.3}$$

Suppose that we require a significance level of 5% to reject the null hypothesis. With a large number of observations, this requires a $t(\hat{\alpha})$ value of at least 1.96. With $\hat{\alpha} = .2$ and $\hat{\sigma}(e) = 2$ we solve Equation 8.3 for *N* and find that

$$1.96 = \frac{.2\sqrt{N}}{2}$$
$$N = 384 \text{ months}$$

or 32 years!

What have we shown? Here is an analyst who has very substantial ability. The example is biased in his favor in the sense that we have assumed away statistical complications. Nothing changes in the parameters over a long period of time. Furthermore, the sample period "behaves" perfectly. Regression estimates are all perfect. Still, it will take Joe's entire working career to get to the point where statistics will confirm his true ability. We have to conclude that even moderate levels of statistical noise make performance evaluation extremely difficult in practice.

Now add to the imprecision of performance estimates the fact that the typical tenure of a fund manager is generally less than 5 years. By the time you are lucky enough to find a fund whose historic superior performance you are confident of, its manager is likely ready to move, or has already moved elsewhere. The nearby box explores this topic further.

Words from the Street Should You Follow Your Fund Manager?

The whole idea of investing in a mutual fund is to leave the stock and bond picking to the professionals. But frequently, events don't turn out quite as expected—the manager resigns, gets transferred, or dies. A big part of the investor's decision to buy a managed fund is based on the manager's record, so changes like these can come as an unsettling surprise. However, a manager's real contribution to fund performance is highly overrated.

Funds are promoted on their managers' track records, which normally span a three- to five-year period. But performance data that goes back only a few years is hardly a valid measure of talent. To be statistically sound, evidence of a manager's track record needs to span, at a minimum, 10 years or more.

The mutual fund industry may look like a merry-go-round of managers, but that shouldn't worry most investors. Many mutual funds are designed to go through little or no change when a manager leaves. That is because, according to a strategy designed to reduce volatility and succession worries, mutual funds are managed by teams of stock pickers, who each run a portion of the assets, rather than by a solo manager with co-captains. Meanwhile, even so-called star managers are nearly always surrounded by researchers and analysts, who can play as much of a role in performance as the manager who gets the headlines.

Don't underestimate the breadth and depth of a fund company's "managerial bench." The larger, established investment companies generally have a large pool of talent to draw on. They are also well aware that investors are prone to depart from a fund when a managerial change occurs.

Lastly, for investors who worry about management changes, there is a solution: index funds. These mutual funds buy stocks and bonds that track a benchmark index like the S&P 500 rather than relying on star managers to actively pick securities. In this case, it doesn't really matter if the manager leaves. Most importantly, index fund investors are not charged the steep fees that are needed to pay star management salaries.

Source: Shauna Carther, "Should You Follow Your Fund Manager?" *Investopedia.com*, March 3, 2010. Provided by *Forbes.*

Concept Check 8.3 Suppose an analyst has a measured alpha of .2% with a standard error of 2%, as in our example. What is the probability that the positive alpha is due to luck of the draw and that true ability is zero?

Selection Bias and Portfolio Evaluation

A warning: Regardless of the performance criterion, some funds will outperform their benchmarks in any year and some will underperform. The good performers are likely to attract more interest from the financial press and potential investors. But beware of focusing on these above-average performers and extrapolating track records into the future. Performance in one period is not predictive of future performance.

When we address the performance of mutual funds selected *because* they have been successful, we need to be even more cautious in evaluating their track records. In particular, we need to recognize that even if all managers were equally skilled, a few "winners" would emerge by sheer chance each period. With thousands of funds in operation, the best-performing funds will have been wildly successful, even if these results reflect luck rather than skill.

Another manifestation of this sort of selection bias (i.e., a focus on nonrepresentative funds) arises when we limit a sample of funds to those for which returns are available over an entire sample period. This practice implies that we exclude from consideration all funds that were closed down over the sample period. The ensuing bias is called **survivorship bias.** It turns out that when even a small number of funds have failed, the upward bias in the performance of surviving funds can be substantial. Most mutual fund databases now include failed funds so that samples can be protected from survivorship bias.

8.2 STYLE ANALYSIS

The index model regression can be viewed as a way to measure and describe facets of a portfolio manager's investment style. Portfolios with high betas are called cyclical or aggressive because they are more responsive to economywide developments. Low beta portfolios are described as defensive. Multifactor models generalize this idea, describing the portfolio's exposure to several risk factors or segments of the market. Each of these exposures can be viewed as an implicit sort of asset allocation decision.

Style analysis was introduced by William Sharpe as a tool to systematically measure the exposures of managed portfolios.[9] The popularity of the concept was aided by a well-known study[10] concluding that most of the variation in returns of 82 mutual funds could be explained by the funds' asset allocation to bills, bonds, and stocks. While later studies have taken issue with the exact interpretation of these results, there is widespread agreement that asset allocation is responsible for a high proportion of the variation across funds in investment performance.

Sharpe's idea was to regress fund returns on indexes representing a range of asset classes. The regression coefficient on each index would then measure the fund's implicit allocation to that "style." Because funds are barred from short positions, the regression coefficients are constrained to be either zero or positive and to sum to 100%, so as to represent a complete asset allocation. The *R*-square of the regression would then measure the percentage of return variability attributable to style choice rather than security selection. The intercept measures the average return from security selection of the fund portfolio. It therefore tracks the average success of security selection over the sample period.

To illustrate Sharpe's approach, we use monthly returns on Fidelity's Magellan Fund during the famous manager Peter Lynch's tenure between October 1986 and September 1991, with results shown in Table 8.4. While seven asset classes are included in this analysis (of which six are represented by stock indexes and one is the T-bill alternative), the regression coefficients are positive for only three, namely, large capitalization stocks, medium cap stocks, and high P/E (growth) stocks. These portfolios alone explain 97.5% of the variance of Magellan's returns. In other words, a tracking portfolio made up of the three style portfolios, with weights as given in Table 8.4, would explain the vast majority of Magellan's variation in monthly performance. We conclude that the fund returns are well represented by three style portfolios.

The proportion of return variability *not* explained by asset allocation can be attributed to security selection within asset classes, as well as timing that shows up as periodic changes in allocation.

Table 8.4 Style Analysis for Fidelity's Magellan Fund

Style Portfolio	Regression Coefficient
T-bill	0
Small cap	0
Medium cap	35
Large cap	61
High P/E (growth)	5
Medium P/E	0
Low P/E (value)	0
Total	100
R-square	97.5

Source: Authors' calculations. Return data for Magellan obtained from **finance.yahoo.com/funds** and return data for style portfolios obtained from the Web page of Professor Kenneth French: **mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.**

For Magellan, residual variability was 100 − 97.5 = 2.5%. This sort of result is commonly interpreted as evidence against the importance of security selection, but such a conclusion misses the important role of the intercept in this regression. (The *R*-square of the regression can be 100%, and yet the intercept can be positive due to consistently superior stock selection.) For Magellan, the intercept was 32 basis points per month, providing a cumulative abnormal return over the 5-year period of 19.19%. The superior performance of Magellan in this period is displayed in Figure 8.4, which plots the combined impact of the intercept plus monthly residuals relative to the tracking portfolio

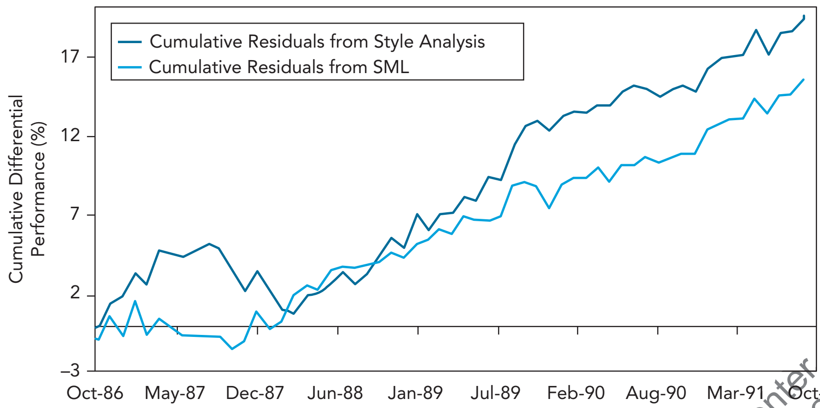


Figure 8.4 **Fidelity Magellan Fund cumulative return difference: Fund versus style benchmark and fund versus SML benchmark.**

Source: Authors' calculations.

[9] William F. Sharpe, "Asset Allocation: Management Style and Performance Evaluation," *Journal of Portfolio Management*, Winter 1992, pp. 7–19.

[10] Gary Brinson, Brian Singer, and Gilbert Beebower, "Determinants of Portfolio Performance," *Financial Analysts Journal*, May/June 1991.

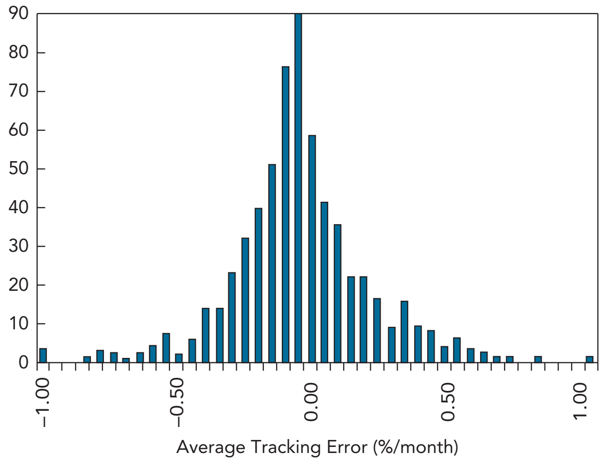


Figure 8.5 **Average tracking error for 636 mutual funds, 1985–1989.**

Source: William F. Sharpe, "Asset Allocation: Management Style and Performance Evaluation," *Journal of Portfolio Management*, Winter 1992, pp. 7–19.

composed of the individual style portfolios, as well as relative to the CAPM benchmark provided by the SML.

Of course, Magellan's consistently positive residual return (reflected in the steadily increasing plot of cumulative return difference) is hardly common. Figure 8.5 shows the frequency distribution of average residuals across 636 mutual funds. The distribution has the familiar bell shape, but with a slightly negative mean of −.074% per month.

Style analysis has become very popular in the investment management industry and has spawned quite a few variations on Sharpe's methodology. Several Web sites can help investors identify managers' style and stock selection performance.

8.3 PERFORMANCE MEASUREMENT WITH CHANGING PORTFOLIO COMPOSITION

One potential problem with risk-adjustment techniques is that they all assume that portfolio risk, whether measured by standard deviation or beta, is constant over the relevant time period. This isn't necessarily so. If a manager attempts to increase portfolio beta when she thinks the market is about to go up and to decrease beta when she's pessimistic, both the standard deviation and the beta of the portfolio will change over time. This can wreak havoc with our performance measures, as illustrated by the following example.

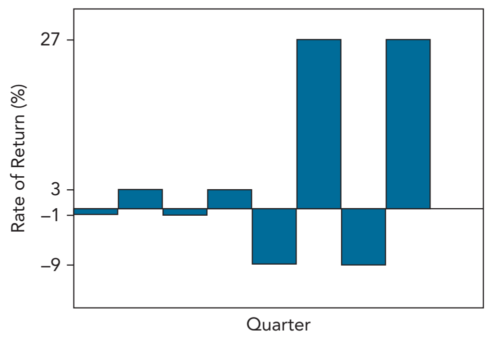


Figure 8.6 **Portfolio returns: Returns in last four quarters are more variable than in the first four.**

Example 8.3 Changing Portfolio Risk

Suppose that the Sharpe measure of the market index is .4. In the first year, the portfolio manager executes a low-risk strategy and realizes an (annualized) mean excess return of 1% and standard deviation of 2%. This makes for a Sharpe ratio of .5, which beats the passive strategy. Over the next year, the manager decides that a *high*-risk strategy is optimal and achieves an annual mean excess return of 9% and standard deviation of 18%. Here, again, the Sharpe ratio is .5. Over the 2-year period, our manager consistently maintains a better-than-passive Sharpe measure.

Figure 8.6 shows a pattern of (annualized) quarterly returns that are consistent with our description of the manager's strategy of two years. In the first four quarters, the excess returns are −1%, 3%, −1%, and 3%, making for an average of 1% and standard deviation of 2%. In the next four quarters, the excess returns are −9%, 27%, −9%, and 27%, making for an average of 9% and standard deviation of 18%. Thus, *both* years exhibit a Sharpe measure of .5. However, the mean and standard deviation of the eight quarterly returns are 5% and 13.42%, respectively, making for a Sharpe measure of only .37, apparently inferior to the passive strategy!

What happened in Example 8.3? The shift of the mean from the first four quarters to the next was not recognized as a shift in strategy. Instead, the difference in mean returns in the two years added to the *appearance* of volatility in portfolio returns. The active strategy with shifting means appears riskier than it really is and biases the estimate of the Sharpe measure downward. We conclude that for actively managed portfolios, it is necessary to keep track of portfolio composition and changes in portfolio mean and risk. We will see another example of this problem in Section 8.4, which deals with market timing.

Performance Manipulation and the Morningstar Risk-Adjusted Rating

We just saw how time-varying risk and return can distort conventional performance evaluation. The problem can be worse when managers, whose compensation depends on performance, try to game the system. Managers observe how returns unfold over the course of the evaluation period and can adjust portfolio strategies (e.g., either increasing or decreasing risk) in an attempt to manipulate performance measures. Once they do so, portfolio strategy in the latter part of the evaluation period comes to depend on performance in the beginning of the period.

Ingersoll, Spiegel, Goetzmann, and Welch[11] show how the conventional performance measures covered in this chapter can be manipulated. While the details of their model are challenging, the logic is straightforward, and we can illustrate using the Sharpe ratio.

Investment in the risk-free asset (lending or borrowing) does not affect the Sharpe ratio of the portfolio. Put differently, the Sharpe ratio is invariant to the fraction y invested in the risky portfolio rather than in the risk-free asset. The reason is that realized excess returns are proportional to y and therefore so are both the risk premium and standard deviation, leaving the Sharpe ratio unchanged. But what if y changes during a period?

Imagine a manager already partway into an evaluation period. While realized excess returns (average return, SD, and Sharpe ratio) are now known for the first part of the evaluation period, the distribution of the remaining future rates is still not determined. The overall Sharpe ratio will be some (complicated) average of the known Sharpe ratio in the first leg and the yet unknown ratio in the second leg of the evaluation period. Increasing leverage during the second leg will increase the influence of the second period on the full-period average performance because leverage will amplify returns, both good and bad. Therefore, managers will wish to increase leverage in the latter part of the period if early returns are poor. Conversely, good first-part performance calls for deleveraging to increase the weight on the initial period. With an extremely good first leg, a manager will shift almost the entire portfolio to the risk-free asset. This strategy induces a (negative) correlation between returns in the first and second legs of the evaluation period.

Investors lose, on average, from this strategy. Arbitrary variation in leverage (and therefore risk) is utility-reducing. It benefits managers only because it allows them to adjust the implicit weighting scheme of the two subperiods over the full evaluation period after observing their initial performance. Hence, investors would like to prohibit or at least eliminate the incentive to pursue this strategy.

Unfortunately, as Ibbotson et al. show, only one performance measure is impossible to manipulate.[12] This is the Morningstar Risk-Adjusted Rating (MRAR). Amazingly, Morning star was not even aiming at a manipulation proof performance measure when it developed its MRAR—it was simply attempting to accommodate investors who wanted a performance measure consistent with constant relative risk aversion. Its measure is defined as follows:

Morningstar risk-adjusted return:

$$\text{MRAR}(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1 + r_t}{1 + r_{ft}}\right)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1$$

The rating is a sort of harmonic average of excess returns, where $t = 1, \ldots, T$ are monthly observations, and γ measures investor risk aversion. For mutual funds, Morningstar uses $\gamma = 2$, which is considered a reasonable value for an average retail client.[13] The MRAR can be interpreted as the risk-free equivalent excess return of the portfolio for an investor with risk aversion measured by γ.

The Morningstar RAR method produces results that are similar but not identical to those of the mean/variance-based Sharpe ratio. Figure 8.7 demonstrates the fit between ranking by RAR and by Sharpe ratios from the performance of 1,286 diversified equity funds over the period 1994–1996. Sharpe notes that this period is characterized by high returns that contribute to a good fit.

Morningstar also computes fund returns adjusted for loads as well as a risk measure based on fund performance in its worst years. Morningstar ranks risk-adjusted performance across funds in a similar style group. The peer group for each fund is selected on the basis of the fund's investment universe (e.g., international, growth versus value, fixed-income) as well as portfolio characteristics such as average price-to-book ratio,

[11] Jonathan Ingersoll, Matthew Spiegel, William Goetzmann, and Ivo Welch, "Portfolio Performance Manipulation and Manipulation Proof Performance Measures," *Review of Financial Studies* 20 (2007).

[12] Ibbotson et al.

[13] The MRAR measure is the *certainty-equivalent geometric average excess return* derived from a more sophisticated utility function than the mean-variance function we used in Chapter 6. The utility function is consistent with *constant relative risk aversion (CRRA).* When investors have CRRA, their capital allocation (the fraction of the portfolio placed in risk-free versus risky assets) does not change with wealth. The coefficient of risk aversion is $A = 1 + \gamma$. When $\gamma = 0$ (equivalently, $A = 1$), the utility function is just the geometric average of gross excess returns:

$$\text{MRAR}(0) = \left[\prod_{t=1}^{T}(1 + R_t)\right]^{\frac{12}{T}} - 1$$

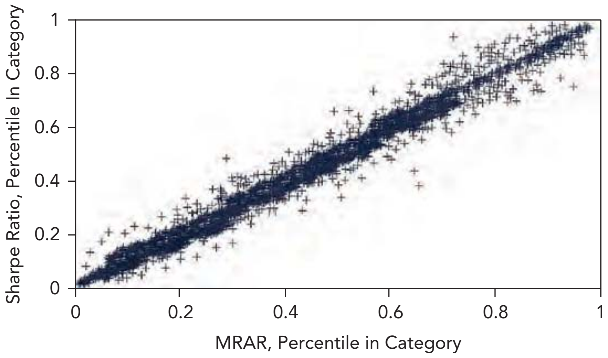


Figure 8.7 **Rankings based on Morningstar's RAR versus Sharpe ratio.**

Source: William F. Sharpe (1997). "Morningstar Performance Measures," **www.standard.edu/-wfsharpe/art/stars/stars0.htm.**

price–earnings ratio, and market capitalization, and stars are awarded based on the following table:

Percentile	Stars
0–10	1
10–32.5	2
32.5–67.5	3
67.5–90	4
90–100	5

The Morningstar five-star rating is coveted by the managers of the thousands of funds covered by the service.

8.4 MARKET TIMING

Another source of variation in portfolio risk is **market timing**. In its pure form, market timing involves shifting funds between a market-index portfolio and a safe asset, depending on whether the index is expected to outperform the safe asset. In practice, most managers do not shift fully between T-bills and the market. How can we account for partial shifts into the market when it is expected to perform well?

To simplify, suppose that an investor holds only the market-index portfolio and T-bills. If the weight of the market were constant, say, .6, then portfolio beta would also be constant, and the SCL would plot as a straight line with slope .6, as in Figure 8.8, Panel A. If, however, the investor could correctly time the market and shift funds into it in periods when the market does well, the SCL would plot as in Figure 8.8, Panel B. If bull and bear markets can be predicted, the investor will shift more into the market when the market is more likely to go up. The portfolio beta and the slope of the SCL will be higher when r_M is higher, resulting in the curved line that appears in Figure 8.8, Panel B.

Treynor and Mazuy were the first to propose estimating such a line by adding a squared term to the usual linear index model:[14]

$$r_P - r_f = a + b(r_M - r_f) + c(r_M - r_f)^2 + e_P$$

where r_P is the portfolio return and a, b, and c are estimated by regression analysis. If c turns out to be positive, we have evidence of timing ability because this last term will make the characteristic line steeper as $r_M - r_f$ is larger. Treynor and Mazuy estimated this equation for a number of mutual funds, but found little evidence of timing ability.

A similar but simpler methodology was proposed by Henriksson and Merton.[15] These authors suggested that the beta of the portfolio takes only two values: a large value if the market is expected to do well and a small value otherwise. Under this scheme, the portfolio characteristic line appears as shown in Figure 8.8, Panel C. Such a line appears in regression form as

$$r_P - r_f = a + b(r_M - r_f) + c(r_M - r_f)D + e_P$$

where D is a dummy variable that equals 1 when $r_M > r_f$ and zero otherwise. Hence the beta of the portfolio is b in bear markets and $b + c$ in bull markets. Again, a positive value of c implies market timing ability. They also found little evidence of market timing ability.

To illustrate how you might implement a test for market timing, return to Table 8.2, which contains 12 months of excess returns for two managed portfolios, P and Q, and the market index, M. Regress the excess returns of each portfolio on the excess returns of M and the square of these returns as in the following specifications:

$$r_P - r_f = a_P + b_P(r_M - r_f) + c_P(r_M - r_f)^2 + e_P$$

$$r_Q - r_f = a_Q + b_Q(r_M - r_f) + c_Q(r_M - r_f)^2 + e_Q$$

You will derive the following statistics. The numbers in parentheses are included for comparison: They are the regression estimates from the single-variable regression reported in Table 8.3.

	Portfolio	
Estimate	***P***	**Q**
Alpha (*a*)	1.77 (1.63)	−2.29 (5.26)
Beta (*b*)	0.70 (0.70)	1.10 (1.40)
Timing (*c*)	0.00	0.10
R-square	0.91 (0.91)	0.98 (0.64)

[14] Jack L. Treynor and Kay Mazuy, "Can Mutual Funds Outguess the Market?" *Harvard Business Review* 43 (July–August 1966).

[15] Roy D. Henriksson and R. C. Merton, "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecast Skills," *Journal of Business* 54 (October 1981).

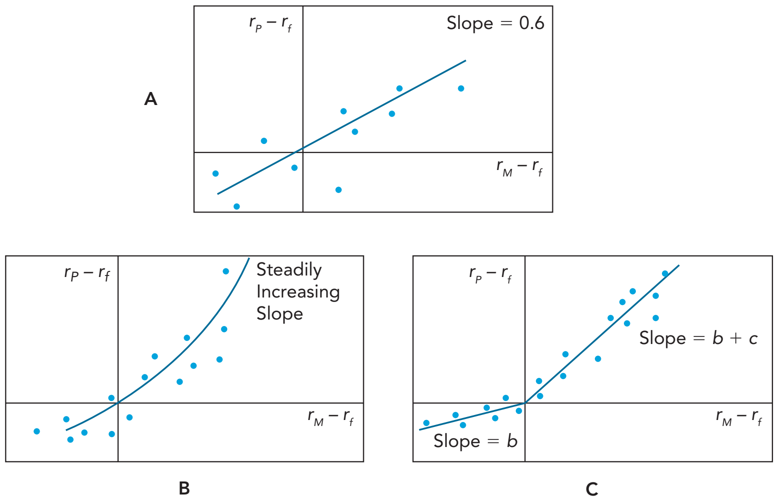


Figure 8.8 **Characteristic lines. *Panel A:* No market timing, beta is constant. *Panel B:* Market timing, beta increases with expected market excess return. *Panel C:* Market timing with only two values of beta.**

Portfolio *P* shows no evidence of attempted timing: Its timing coefficient, *c*, is estimated to be zero. It is not clear whether this is because no attempt was made at timing or because any effort to time the market was in vain and served only to increase portfolio variance unnecessarily.

The results for portfolio Q, however, reveal that timing has, in all likelihood, been attempted. Here the coefficient, *c*, is positive, with an estimated value of .10. The evidence thus suggests successful timing, offset by unsuccessful stock selection (negative *a*). Note that the estimate of alpha, *a*, is now −2.29% as opposed to the 5.26% estimate derived from the regression equation that did not account for the potential impact of timing.

This example illustrates the inadequacy of conventional performance evaluation techniques that assume constant mean returns and constant risk. The market timer constantly shifts beta and mean return, moving into and out of the market. So market timing presents another instance in which portfolio composition and risk change over time, complicating the effort to evaluate performance. Whereas the expanded regression captures this possibility, the simple SCL does not. The important point for performance evaluation is that expanded regressions can capture many of the effects of portfolio composition change that would confound more conventional mean-variance measures.

The Potential Value of Market Timing

Suppose we define perfect market timing as the ability to tell (with certainty) at the beginning of each year whether the S&P 500 portfolio will outperform Treasury bills. Accordingly, at the beginning of each year, the market timer shifts all funds into either cash equivalents (T-bills) or equities (the S&P portfolio), whichever is predicted to do better. Beginning with $1 on December 31, 1926, how would the perfect timer end an 92-year experiment on December 31, 2018, in comparison with investors who kept their funds in either equity or T-bills for the entire period?

Table 8.5 presents summary statistics for each of the three passive strategies, computed from the historical annual returns of bills and equities. From the returns on stocks and bills, we calculate wealth indexes of the all-bills and all-equity investments and show terminal values for these investors at the end of 2018. The return for the perfect timer in each year is the *maximum* of the return on stocks and the return on bills.

The first row in Table 8.5 shows the terminal value of investing $1 in bills over the 92 years (1926–2018) is $20, while the terminal value of the same initial investment in equities is $5,271. That as impressive as the difference in terminal values is, it is best interpreted as no more than compensation for the risk borne by equity investors. As we've already seen, the *annual* difference in returns is just about 8%, which doesn't seem as dramatic. Notice that the standard deviation of the

Table 8.5 **Performance of Bills, Equities, and Perfect (Annual) Market Timers. Initial Investment = $1**

Strategy	Bills	Equities	Perfect Timer
Terminal value	$20	$5,271	$755,809
Arithmetic average	3.39%	11.49%	16.41%
Standard deviation	3.14%	20.04%	13.44%
Geometric average	3.34%	9.76%	15.85%
Maximum	14.71%	57.35%	57.35%
Minimum*	−0.02%	−44.04%	0.00%
Skew	1.05	−0.39	0.75
Kurtosis	1.01	0.07	−0.07
LPSD	0.00%	13.10%	0.00%

* A negative rate on "bills" was observed in 1940. The Treasury security used in the data series in these early years was actually not a T-bill but a T-bond with 30 days to maturity.

all-equity investor was a hefty 20.04%. This is also why the geometric average of stocks for the period is "only" 9.76%, compared with the arithmetic average of 11.49%. (The difference between the two averages increases with volatility.)

Now observe that the terminal value of the perfect timer is $755,809, a 143-fold increase over the already large terminal value of the all-equity strategy! In fact, this result is even better than it looks because the return to the market timer is truly risk-free. This is the classic case where a large standard deviation (13.44%) has nothing to do with risk. Because the timer never delivers a return below the risk-free rate, the standard deviation is a measure of *good* surprises only.

The positive skew of the timer's distribution (.75, compared with the negative skew of equities) is a manifestation of the fact that the extreme values are all positive. Other indications of this stellar performance are the minimum and maximum returns—the minimum return for the timer equals zero (in 1940) and the maximum return, 57.35%, is that of equities (in 1933). All negative returns on equities (as low as −44% in 1931) were avoided by the timer. Finally, another clear indication of the advantage of the perfect timer compared to the all-equity portfolio is provided by the lower partial standard deviation, LPSD.[16] The LPSD, which is a measure of the typical amplitude of any return shortfall compared to T-bills, is 13.10% for the all-equity portfolio, but it is necessarily zero for the perfect timer (who always performs *at least* as well as bills).

Valuing Market Timing as a Call Option

The key to valuing market timing ability is to recognize that perfect foresight is equivalent to holding a call option on the equity portfolio—but without having to pay for it! The perfect timer invests 100% in either the safe asset or the equity portfolio, whichever will provide the higher return. The rate of return is atleast the risk-free rate. This is shown in Figure 8.9.

To see how timing skill can be viewed as a free option, suppose that the market index currently is at S_0 and that a call option on the index has an exercise price of $X = S_0(1 + r_f)$. If the market outperforms bills over the coming period, S_T will exceed X; otherwise it will be less than X. Now look at the payoff to a portfolio consisting of this option and S_0 dollars invested in bills:

[16] The LPSD is sometimes based on the average squared deviation below the mean. Because the threshold performance in this application is the risk-free rate, we calculate the LPSD for this discussion by taking squared deviations from that rate, and the observations are conditional on being below that threshold. It ignores the number of such events.

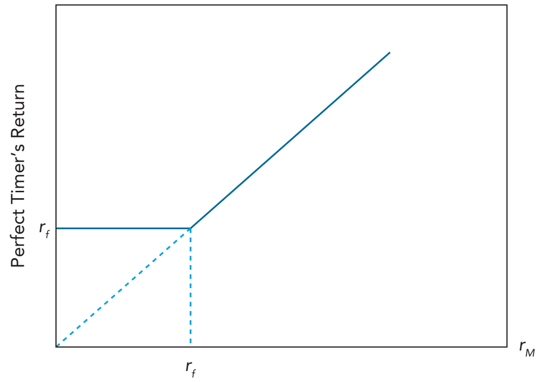


Figure 8.9 **Rate of return of a perfect market timer as a function of the rate of return on the market index.**

	$S_T < X$	$S_T \geq X$
Bills	$S_0(1 + r_f)$	$S_0(1 + r_f)$
Call	0	$S_T - X$
Total	$S_0(1 + r_f)$	S_T

The portfolio pays the risk-free return when the market is bearish (i.e., the market return is less than the risk-free rate), and it pays the market return when the market is bullish and outperforms bills. Such a portfolio is a perfect market timer.[17]

Because the ability to predict the better-performing investment is equivalent to acquiring a (free) call option on the market and adding it to a position in bills, we can use option-pricing models to assign a dollar value to perfect timing ability. This value would constitute the fair fee that a perfect timer could charge investors for its services. Placing a value on perfect timing also enables us to assign value to less-than-perfect timers.

The exercise price of the perfect-timer call option on \$1 of the equity portfolio is the final value of the T-bill investment. Using continuous compounding, this is \$1 $\times\, e^{rT}$. When you use this exercise price in the Black-Scholes formula for the value of the call option, the formula simplifies considerably to[18]

$$\text{MV(Perfect timer per \$ of assets)} =$$
$$C = 2N(\tfrac{1}{2}\,\sigma_M\sqrt{T}) - 1 \qquad (8.4)$$

[17] The analogy between market timing and call options, and the valuation formulas that follow from it, were developed in Robert C. Merton, "On Market Timing and Investment Performance: An Equilibrium Theory of Value for Market Forecasts," *Journal of Business*, July 1981.

[18] Substitute $S_0 = \$1$ for the current value of the equity portfolio and $X = \$1 \times e^{rT}$ in Equation 21.1 of Chapter 21, and you will obtain Equation 8.4.

We have so far assumed annual forecasts, that is, $T = 1$ year. Using $T = 1$, and the standard deviation of stocks from Table 8.5, 20.04%, we compute the value of this call option as 7.98 cents, or 7.98% of the value of the equity portfolio.[19]

The Value of Imperfect Forecasting

A weather forecaster in Tucson, Arizona, who *always* predicts no rain may be right 90% of the time. But a high success rate for a "stopped-clock" strategy is not evidence of forecasting ability. Similarly, the appropriate measure of market forecasting ability is not the overall proportion of correct forecasts. If the market is up two days out of three and a forecaster *always* predicts market advance, the resulting two-thirds success rate does not imply forecasting ability. We need to examine the proportion of bull markets ($r_M < r_f$) correctly forecast *and* the proportion of bear markets ($r_M > r_f$) correctly forecast.

If we call P_1 the proportion of the correct forecasts of bull markets and P_2 the proportion for bear markets, then $P_1 + P_2 - 1$ is the correct measure of timing ability. For example, a forecaster who always guesses correctly will have $P_1 = P_2 = 1$, and will show ability of $P_1 + P_2 - 1 = 1$ (100%). An analyst who always bets on a bear market will mispredict all bull markets ($P_1 = 0$), will correctly "predict" all bear markets ($P_2 = 1$), and will end up with timing ability of $P_1 + P_2 - 1 = 0$.

Concept Check 8.4 What is the market timing score of someone who flips a fair coin to predict the market?

Merton shows that the value of imperfect market timing is equal to a portion of a call option. The value of an imperfect timer is simply the value of the perfect-timing call times our measure of timing ability, $P_1 + P_2 - 1$.[20]

$$\text{MV(Imperfect timer)} = (P_1 + P_2 - 1) \times C = (P_1 + P_2 - 1)\,[2N(\tfrac{1}{2}\sigma_M\sqrt{T}) - 1] \tag{8.5}$$

The incredible potential payoff to accurate timing versus the relative scarcity of billionaires suggests that market timing is far from a trivial exercise and that very imperfect timing is the most that we can hope for.

[19] This is less than the historical-average excess return of perfect timing shown in Table 8.5, reflecting the fact that the actual value of timing is sensitive to fat tails in the distribution of returns, whereas Black-Scholes assumes a log-normal distribution.

[20] Notice that Equation 8.5 implies that an investor with a value of $P = 0$ who attempts to time the market would add zero value. The shifts across markets would be no better than a random decision concerning asset allocation.

8.5 PERFORMANCE ATTRIBUTION PROCEDURES

Rather than focus on risk-adjusted returns, practitioners often want simply to ascertain which decisions resulted in superior or inferior performance. Superior investment performance depends on an ability to be in the "right" securities at the right time. Such timing and selection ability may be considered broadly, such as being in equities as opposed to fixed-income securities when the stock market is performing well. Or it may be defined at a more detailed level, such as choosing the relatively better-performing stocks within a particular industry.

Portfolio managers continually make broad-brush asset allocation decisions as well as more detailed sector and security allocation decisions within asset classes. Performance attribution studies attempt to decompose overall performance into discrete components that may be identified with a particular level of the portfolio selection process.

Attribution analysis starts from the broadest asset allocation choices and progressively focuses on ever-finer details of portfolio choice. The difference between a managed portfolio's performance and that of a benchmark portfolio then may be expressed as the sum of the contributions to performance of a series of decisions made at the various levels of the portfolio construction process. For example, one common attribution system decomposes performance into three components: (1) broad asset market allocation choices across equity, fixed-income, and money markets; (2) industry (sector) choice within each market; and (3) security choice within each sector.

The attribution method explains the difference in returns between a managed portfolio, *P*, and a selected benchmark portfolio, *B*, called the **bogey.** The bogey is designed to measure the returns the portfolio manager would earn if he or she were to follow a completely passive strategy. "Passive" in this context has two attributes. First, it means that the allocation of funds across broad asset classes is set in accord with a notion of "usual," or neutral, allocation across sectors. This would be considered a passive asset-market allocation. Second, it means that *within* each asset class, the portfolio manager holds an indexed portfolio, such as the S&P 500 index for the equity sector. In such a manner, the passive strategy used as a performance benchmark rules out asset allocation as well as security selection decisions. Any departure of the manager's return from the passive benchmark must be due to either asset allocation bets (departures from the neutral allocation across markets) or security selection bets (departures from the passive index within asset classes).

It is worth briefly explaining the determination of the neutral allocation of funds across the broad asset classes. Weights that

are designated as "neutral" will depend on the risk tolerance of the investor and must be determined in consultation with the client. For example, risk-tolerant clients may place a large fraction of their portfolio in the equity market, perhaps directing the fund manager to set neutral weights of 75% equity, 15% bonds, and 10% cash equivalents. Any deviation from these weights must be justified by a belief that one or another market will either over-or underperform its usual risk-return profile. In contrast, more risk-averse clients may set neutral weights of 45%/35%/20% for the three markets. Therefore, their portfolios in normal circumstances will be exposed to less risk than that of the risk-tolerant client. Only intentional bets on market performance will result in departures from this profile.

To illustrate, consider the attribution results for a hypothetical portfolio. The portfolio invests in stocks, bonds, and money market securities. An attribution analysis appears in Tables 8.6 through 8.9. The portfolio return over the month is 5.34%.

In Table 8.6, the neutral weights have been set at 60% equity, 30% fixed income, and 10% cash (money market securities). The bogey portfolio, comprised of investments in each index with the 60/30/10 weights, returned 3.97%. The managed portfolio's measure of performance is positive and equal to its actual return less the return of the bogey: 5.34 − 3.97 = 1.37%. The next step is to allocate the 1.37% excess return to the separate decisions that contributed to it.

Asset Allocation Decisions

Table 8.7 shows that in this month, the manager established asset allocation weights of 70% in equity, 7% in fixed income, and 23% in cash equivalents. To isolate the effect of this departure from neutral asset allocation, we compare the performance of a hypothetical portfolio that would have been invested in a passive index for each market with weights 70/7/23 versus one invested in each index using the benchmark 60/30/10 weights. This return difference measures the effect of the shift away from the benchmark weights without allowing for any effects attributable to active management of the securities selected within each market.

Superior performance relative to the bogey is achieved by overweighting investments in markets that turn out to perform well and by underweighting those in poorly performing markets. The contribution of asset allocation to superior performance equals the sum over all markets of the excess weight (sometimes called the *active weight* in the industry) in each market times the return of the index for that market.

Panel A of Table 8.7 demonstrates that asset allocation contributed 31 basis points to the portfolio's overall excess return of 137 basis points. The major factor contributing to superior performance in this month is the heavy weighting of the equity market in a month when the equity market has an excellent return of 5.81%.

Sector and Security Selection Decisions

If .31% of the excess performance (Table 8.7, Panel A) can be attributed to advantageous asset allocation *across* markets, the remaining 1.06% must be attributable to sector selection and security selection *within* each market. Table 8.7, Panel B, details the contribution of the managed portfolio's sector and security selection to total performance.

Panel B shows that the equity component of the managed portfolio has a return of 7.28% versus a return of 5.81% for the S&P 500. The fixed-income return is 1.89% versus 1.45%

Table 8.6 Performance of the Managed Portfolio

	Bogey Performance and Excess Return	
Component	**Benchmark Weight**	**Return of Index during Month (%)**
Equity (S&P 500)	0.60	5.81
Bonds (Barclays Aggregate Index)	0.30	1.45
Cash (money market)	0.10	0.48
Bogey = (0.60 × 5.81) + (0.30 × 1.45) + (0.10 × 0.48) = 3.97%		
Return of managed portfolio	5.34%	
−Return of bogey portfolio	3.97	
Excess return of managed portfolio	1.37%	

Table 8.7 Performance Attribution

A. Contribution of Asset Allocation to Performance

Market	(1) Actual Weight in Market	(2) Benchmark Weight in Market	(3) Active or Excess Weight	(4) Index Return (%)	(5) = (3) × (4) Contribution to Performance (%)
Equity	0.70	0.60	0.10	5.81	0.5810
Fixed-income	0.07	0.30	−0.23	1.45	−0.3335
Cash	0.23	0.10	0.13	0.48	0.0624
Contribution of asset allocation					0.3099

B. Contribution of Selection to Total Performance

Market	(1) Portfolio Performance (%)	(2) Index Performance (%)	(3) Excess Performance (%)	(4) Portfolio Weight	(5) = (3) × (4) Contribution (%)
Equity	7.28	5.81	1.47	0.70	1.03
Fixed-income	1.89	1.45	0.44	0.07	0.03
Contribution of selection within markets					1.06

Table 8.8 Sector Selection within the Equity Market

	(1)	(2)	(3)	(4)	(5) = (3) × (4)
	Beginning of Month Weights (%)				
Sector	Portfolio	S&P 500	Active Weight (%)	Sector Return (%)	Sector Allocation Contribution
Basic materials	1.96	8.3	−6.34	6.9	−0.4375
Business services	7.84	4.1	3.74	7.0	0.2618
Capital goods	1.87	7.8	−5.93	4.1	−0.2431
Consumer cyclical	8.47	12.5	−4.03	8.8	0.3546
Consumer noncyclical	40.37	20.4	19.97	10.0	1.9970
Credit sensitive	24.01	21.8	2.21	5.0	0.1105
Energy	13.53	14.2	−0.67	2.6	−0.0174
Technology	1.95	10.9	−8.95	0.3	−0.0269
Total					1.2898

for the Barclays Aggregate Bond Index. The superior performance in both equity and fixed-income markets weighted by the portfolio proportions invested in each market sums to the 1.06% contribution to performance attributable to sector and security selection.

Table 8.8 documents the decisions that led to the superior equity market performance. The first three columns detail the allocation of funds within the equity market compared to their representation in the S&P 500. Column (4) shows the rate of return of each sector. The contribution of each sector's

EXCEL APPLICATIONS Performance Attribution

The performance attribution spreadsheet develops the attribution analysis presented in this section. The model can be used to analyze the performance of mutual funds and other managed portfolios.

You can find this Excel model in Connect.

Excel Questions

1. What would happen to the contribution of asset allocation to overall performance if the actual weights had been 75/12/13 instead of 70/7/23? Explain your result.
2. What would happen to the contribution of security selection to overall performance if the actual return on the equity portfolio had been 6.81% instead of 5.81% and the return on the bond portfolio had been 0.45% instead of 1.45%? Explain your result.

	A	B	C	D	E	F
1	Performance Attribution					
2						
3						
4	Bogey					
5	Portfolio		Benchmark	Return on	Portfolio	
6	Component	Index	Weight	Index	Return	
7	Equity	S&P 500	0.60	5.8100%	3.4860%	
8	Bonds	Barclays Index	0.30	1.4500%	0.4350%	
9	Cash	Money Market	0.10	0.4800%	0.0480%	
10			Return on Bogey		3.9690%	
11						
12		Managed				
13		Portfolio	Portfolio	Actual	Portfolio	
14		Component	Weight	Return	Return	
15		Equity	0.70	5.8100%	5.0960%	
16		Bonds	0.07	1.4500%	0.1323%	
17		Cash	0.23	0.4800%	0.1104%	
18			Return on Managed		5.3387%	
19			Excess Return		1.3697%	

allocation presented in column (5) equals the product of the difference in the sector weight and the sector's performance.

Good performance (a positive contribution) derives from overweighting well performing sectors such as consumer non-cyclicals. The excess return of the equity component of the portfolio attributable to sector allocation alone is 1.29%. Table 8.7, Panel B, column (3), shows that the equity component of the portfolio outperformed the S&P 500 by 1.47%. We conclude that the effect of security selection *within* sectors must have contributed an additional 1.47% −1.29%, or .18%, to the performance of the equity component of the portfolio.

A similar sector analysis can be applied to the fixed-income portion of the portfolio, but we do not show those results here.

Summing Up Component Contributions

In this particular month, all facets of the portfolio selection process were successful. Table 8.9 details the contribution of each aspect of performance. Asset allocation across the major security markets contributes 31 basis points. Sector and security allocation within those markets contributes 106 basis points, for total excess portfolio performance of 137 basis points.

The sector and security allocation of 106 basis points can be partitioned further. Sector allocation within the equity market results in excess performance of 129 basis points, and security selection within sectors contributes 18 basis points. (The total equity excess performance of 147 basis points is multiplied by the 70% weight in equity to obtain contribution to portfolio performance.) Similar partitioning could be done for the fixed-income sector.

Table 8.9 Portfolio Attribution: Summary

		Contribution (basis points)
1. Asset allocation		31
2. Selection		
a. Equity excess return (basis points)		
i. Sector allocation	129	
ii. Security selection	18	
	147 × 0.70 (portfolio weight) =	102.9
b. Fixed-income excess return	44 × 0.07 (portfolio weight) =	3.1
Total excess return of portfolio		137.0

Concept Check 8.5

a. Suppose the benchmark weights in Table 8.7 had been set at 70% equity, 25% fixed-income, and 5% cash equivalents. What would have been the contributions of the manager's asset allocation choices?

b. Suppose the S&P 500 return is 5%. Compute the new value of the manager's security selection choices.

SUMMARY

1. The simplest performance measure compares average return to that on a benchmark such as an appropriate market index or even the median return of funds in a comparison universe. Alternative measures of the average return include the arithmetic and geometric average and time-weighted versus dollar-weighted returns.

2. The appropriate performance measure depends on the role of the portfolio to be evaluated. Appropriate performance measures are as follows:

   a. Sharpe: When the portfolio represents the entire investment fund.

   b. Information ratio: When the portfolio represents the active portfolio to be optimally mixed with the passive portfolio.

   c. Treynor: When the portfolio represents one subportfolio of many.

   d. Jensen (alpha): All of these measures require a positive alpha for the portfolio to be considered attractive.

3. Many observations and long sample periods are required to eliminate the effect of the "luck of the draw" from the evaluation process because portfolio returns commonly are very "noisy."

4. Style analysis uses a multiple regression model where the factors are category (style) portfolios such as bills, bonds, and stocks. The coefficients on the style portfolios indicate a passive strategy that would match the risk exposures of the managed portfolio. These differences in returns between the managed portfolio and the matching portfolio measure performance relative to similar-style funds.

5. Shifting mean and risk of actively managed portfolios make it difficult to assess performance. An important example of this problem arises when portfolio managers attempt to time the market, resulting in ever-changing portfolio betas.

6. One way to measure timing and selection success simultaneously is to estimate an expanded security characteristic line, for which the slope (beta) coefficient is allowed to increase as the market return increases. Another way to evaluate timers is based on the implicit call option embedded in their performance.

7. Common attribution procedures decompose portfolio performance to asset allocation, sector selection, and security selection decisions. Performance is assessed by calculating departures of portfolio composition from a benchmark or neutral portfolio.

Key Terms

time-weighted average
dollar-weighted rate of return
comparison universe
Sharpe ratio
Treynor's measure
Jensen's alpha
information ratio
survivorship bias
market timing
bogey

Key Equations

Geometric time-weighted return: $1 + r_G = [(1 + r_1)(1 + r_2) \ldots (1 + r_n)]^{1/n}$

Sharpe ratio: $S_P = \dfrac{r_P - r_f}{\sigma_P}$

M^2 of portfolio P given its Sharpe ratio: $M^2 = \sigma_M(S_P - S_M)$

Treynor measure: $T_P = \dfrac{r_P - r_f}{\beta_P}$

Jensen's alpha: $\alpha_P = \bar{r}_p - [\bar{r}_f + \beta_p(\bar{r}_M - \bar{r}_f)]$

Information ratio: $\dfrac{\alpha_P}{\sigma(e_P)}$

Morningstar risk-adjusted return:

$$\text{MRAP}(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1 + r_t}{1 + r_{ft}}\right)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1$$

PROBLEM SETS

1. A household savings-account spreadsheet shows the following entries:

Date	Additions	Withdrawals	Value
1/1/2019			148,000
1/3/2019	2,500		
3/20/2019	4,000		
7/5/2019	1,500		
12/2/2019	13,460		
3/10/2020		23,000	
4/7/2020	3,000		
5/3/2020			198,000

Use the Excel function XIRR to calculate the dollar-weighted average return between the first and final dates.

2. Is it possible for a positive alpha to be associated with inferior performance? Explain.

3. We know that the geometric average (time-weighted return) on a risky investment is always lower than the corresponding arithmetic average. Can the IRR (the dollar-weighted return) similarly be ranked relative to these other two averages?

4. We have seen that market timing has tremendous potential value. Would it therefore be wise to shift resources to timing at the expense of security selection?

5. Consider the rate of return of stocks ABC and XYZ.

Year	r_{ABC}	r_{XYZ}
1	20%	30%
2	12	12
3	14	18
4	3	0
5	1	−10

a. Calculate the arithmetic average return on these stocks over the sample period.

b. Which stock has greater dispersion around the mean return?

c. Calculate the geometric average returns of each stock. What do you conclude?

d. If you were equally likely to earn a return of 20%, 12%, 14%, 3%, or 1% in each year (these are the five annual returns for stock ABC), what would be your expected rate of return?

e. What if the five possible outcomes were those of stock XYZ?

f. Given your answers to parts (*d*) and (*e*), which measure of average return, arithmetic or geometric, appears more useful for predicting future performance?

6. XYZ's stock price and dividend history are as follows:

Year	Beginning-of-Year Price	Dividend Paid at Year-End
2018	$100	$4
2019	120	4
2020	90	4
2021	100	4

An investor buys three shares of XYZ at the beginning of 2018, buys another two shares at the beginning of 2019, sells one share at the beginning of 2020, and sells all four remaining shares at the beginning of 2021.

a. What are the arithmetic and geometric average time-weighted rates of return for the investor?

b. What is the dollar-weighted rate of return? (*Hint:* Carefully prepare a chart of cash flows for the *four* dates corresponding to the turns of the year for January 1, 2018, to January 1, 2021. If your calculator cannot calculate internal rate of return, you will have to use trial and error.)

7. A manager buys three shares of stock today and then sells one of those shares each year for the next three years. His actions and the price history of the stock are summarized below. The stock pays no dividends.

Time	Price	Action
0	$ 90	Buy 3 shares
1	100	Sell 1 share
2	100	Sell 1 share
3	100	Sell 1 share

a. Calculate the time-weighted geometric average return on this "portfolio."

b. Calculate the time-weighted arithmetic average return on this portfolio.

c. Calculate the dollar-weighted average return on this portfolio.

8. Based on current dividend yields and expected capital gains, the expected rates of return on portfolios *A* and *B* are 12% and 16%, respectively. The beta of *A* is .7, while that of *B* is 1.4. The T-bill rate is currently 5%, whereas the expected rate of return of the S&P 500 index is 13%. The standard deviation of portfolio *A* is 12% annually, that of *B* is 31%, and that of the S&P 500 index is 18%.

a. If you currently hold a market-index portfolio, would you choose to add either of these portfolios to your holdings? Explain.

b. If instead you could invest *only* in T-bills and *one* of these portfolios, which would you choose?

9. Consider the two (excess return) index-model regression results for stocks *A* and *B*. The risk-free rate over the period was 6%, and the market's average return was 14%. Performance is measured using an index model regression on excess returns.

	Stock A	Stock B
Index model regression estimates	$1\% + 1.2(r_M - r_f)$	$2\% + 0.8(r_M - r_f)$
R-square	0.576	0.436
Residual standard deviation, $\sigma(e)$	10.3%	19.1%
Standard deviation of excess returns	21.6%	24.9%

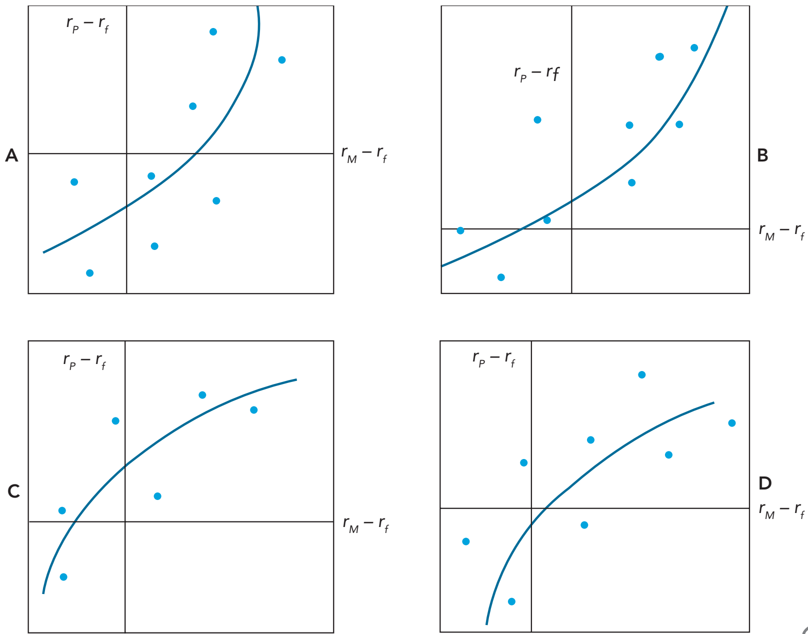
a. Calculate the following statistics for each stock:

i. Alpha
ii. Information ratio
iii. Sharpe ratio
iv. Treynor measure

b. Which stock is the best choice under the following circumstances?

i. This is the only risky asset to be held by the investor.
ii. This stock will be mixed with the rest of the investor's portfolio, currently composed solely of holdings in the market-index fund.
iii. This is one of many stocks that the investor is analyzing to form an actively managed stock portfolio.

10. Evaluate the market timing and security selection abilities of four managers whose performances are plotted in the accompanying diagrams.



11. Consider the following information regarding the performance of a money manager in a recent month. The table represents the actual return of each sector of the manager's portfolio in column 1, the fraction of the portfolio allocated to each sector in column 2, the benchmark or neutral sector allocations in column 3, and the returns of sector indices in column 4.

	Actual Return	Actual Weight	Benchmark Weight	Index Return
Equity	2%	0.70	0.60	2.5% (S&P 500)
Bonds	1	0.20	0.30	1.2 (Barclay's Aggregate)
Cash	0.5	0.10	0.10	0.5

a. What was the manager's return in the month? What was her overperformance or underperformance?

b. What was the contribution of security selection to relative performance?

c. What was the contribution of asset allocation to relative performance? Confirm that the sum of selection and allocation contributions equals her total "excess" return relative to the bogey.

12. A global equity manager is assigned to select stocks from a universe of large stocks throughout the world. The manager will be evaluated by comparing her returns to the return on the MSCI World Market Portfolio, but she is free to hold stocks from various countries in whatever proportions she finds desirable. Results for a given month are contained in the following table:

Country	Weight in MSCI Index	Manager's Weight	Manager's Return in Country	Return of Stock Index for That Country
U.K.	0.15	0.30	20%	12%
Japan	0.30	0.10	15	15
U.S.	0.45	0.40	10	14
Germany	0.10	0.20	5	12

a. Calculate the total value added of all the manager's decisions this period.

b. Calculate the value added (or subtracted) by her *country* allocation decisions.

c. Calculate the value added from her stock selection ability within countries.

d. Confirm that the sum of the contributions to value added from her country allocation plus security selection decisions equals total over- or underperformance.

13. Conventional wisdom says that one should measure a manager's investment performance over an entire market cycle. What arguments support this convention? What arguments contradict it?

14. Does the use of universes of managers with similar investment styles to evaluate relative investment performance overcome the statistical problems associated with instability of beta or total variability?

15. During a particular year, the T-bill rate was 6%, the market return was 14%, and a portfolio manager with beta of .5 realized a return of 10%.

a. Evaluate the manager based on the portfolio alpha.

b. Reconsider your answer to part (*a*) in view of the empirical finding that the security market line is too flat. Does this affect your assessment of performance?

16. Bill Smith is evaluating the performance of four large-cap equity portfolios: Funds *A*, *B*, *C*, and *D*. As part of his analysis, Smith computed the Sharpe ratio and the Treynor measure for all four funds. Based on his finding, the ranks assigned to the four funds are as follows:

Fund	Treynor Measure Rank	Sharpe Ratio Rank
A	1	4
B	2	3
C	3	2
D	4	1

The difference in rankings for Funds *A* and *D* is most likely due to:

a. A lack of diversification in Fund *A* as compared to Fund *D*.

b. Different benchmarks used to evaluate each fund's performance.

c. A difference in risk premiums.

Use the following information to answer Problems 17 through 20: Primo Management Co. is looking at how best to evaluate the performance of its managers. Primo has been hearing more and more about benchmark portfolios and is interested in trying this approach. As such, the company hired Sally Jones, CFA, as a consultant to educate the managers on the best methods for constructing a benchmark portfolio, how to choose the best benchmark, whether the style of the fund under management matters, and what they should do with their global funds in terms of benchmarking.

For the sake of discussion, Jones put together some comparative 2-year performance numbers that relate to Primo's current domestic funds under management and a potential benchmark.

Style Category	Weight		Return	
	Primo	Benchmark	Primo	Benchmark
Large-cap growth	0.60	0.50	17%	16%
Mid-cap growth	0.15	0.40	24	26
Small-cap growth	0.25	0.10	20	18

As part of her analysis, Jones also takes a look at one of Primo's global funds. In this particular portfolio, Primo is invested 75% in Dutch stocks and 25% in British stocks. The benchmark is invested 50% in Dutch stocks and and 50% in British stocks. On average, the British stocks outperformed the Dutch stocks. The euro appreciated 6% versus the U.S. dollar over the holding period while the pound depreciated 2% versus the dollar. In terms of the local return, Primo outperformed the benchmark with the Dutch investments but underperformed the index with respect to the British stocks.

17. What is the within-sector selection effect for each individual sector?

18. Calculate the amount by which the Primo portfolio out-(under-)performed the market over the period, as well as the contribution to performance of the pure sector allocation and security selection decisions.

19. If Primo decides to use return-based style analysis, will the R^2 of the regression equation of a passively managed fund be higher or lower than that of an actively managed fund?

20. Which of the following statements about Primo's global fund is most correct? Primo appears to have a positive currency allocation effect as well as

a. A negative market allocation effect and a positive security allocation effect.

b. A negative market allocation effect and a negative security allocation effect.

c. A positive market allocation effect and a negative security allocation effect.

21. Kelli Blakely is a portfolio manager for the Miranda Fund, a core large-cap equity fund. The market proxy and benchmark for performance measurement purposes is the S&P 500. Although the Miranda portfolio generally mirrors the asset class and sector weightings of the S&P, Blakely is allowed a significant amount of leeway in managing the fund.

Blakely was able to produce exceptional returns last year (as outlined in the table below) through her market timing and security selection skills. At the outset of the year, she became extremely concerned that the combination of a weak economy and geopolitical uncertainties would negatively impact the market. Taking a bold step, she changed her market allocation. For the entire year her asset class exposures averaged 50% in stocks and 50% in cash. The S&P's allocation between stocks and cash during the period was a constant 97% and 3%, respectively. The risk-free rate of return was 2%.

One-Year Trailing Returns		
	Miranda Fund	S&P 500
Return	10.2%	−22.5%
Standard deviation	37%	44%
Beta	1.10	1.00

a. What are the Sharpe ratios for the Miranda Fund and the S&P 500?

b. What are the M^2 measures for the Miranda Fund and the S&P 500?

c. What is the Treynor measure for the Miranda Fund and the S&P 500?

d. What is the Jensen measure for the Miranda Fund?

22. Go to Kenneth French's data library site at **http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html**. Select two industry portfolios of your choice and download 36 months of data. Download other data from the site as needed to perform the following tasks.

a. Compare the portfolio's performance to that of the market index on the basis of the Sharpe, Jensen, Treynor measures as well as the information ratio. Plot the monthly values of alpha plus residual return.

b. Now use the Fama-French three-factor model (see Section 13.3) as the return benchmark. Compute plots of alpha plus residual return using the FF model. How does performance change using this benchmark instead of the market index?

CFA® PROBLEMS

1. You and a prospective client are considering the measurement of investment performance, particularly with respect to international portfolios for the past five years. The data you discussed are presented in the following table:

International Manager or Index	Total Return	Country and Security Return	Currency Return
Manager A	−6.0%	2.0%	−8.0%
Manager B	−2.0	−1.0	−1.0
International Index	−5.0	0.2	−5.2

a. Assume that the data for manager A and manager B accurately reflect their investment skills and that both managers actively manage currency exposure. Briefly describe one strength and one weakness for each manager.

b. Recommend and justify a strategy that would enable your fund to take advantage of the strengths of each of the two managers while minimizing their weaknesses.

2. Carl Karl, a portfolio manager for the Alpine Trust Company, has been responsible since 2023 for the City of Alpine's Employee Retirement Plan, a municipal pension fund. Alpine is a growing community, and city services and employee payrolls have expanded in each of the past 10 years. Contributions to the plan in fiscal 2028 exceeded benefit payments by a three-to-one ratio.

The plan board of trustees directed Karl five years ago to invest for total return over the long term. However, as trustees of this highly visible public fund, they cautioned him that volatile or erratic results could cause them embarrassment. They also noted a state statute that mandated that not more than 25% of the plan's assets (at cost) be invested in common stocks.

At the annual meeting of the trustees in November 2028, Karl presented the following portfolio and performance report to the board:

Alpine Employee Retirement Plan				
Asset Mix as of 9/30/2028	**At Cost (millions)**		**At Market (millions)**	
Fixed-income assets:				
Short-term securities	$ 4.5	11.0%	$ 4.5	11.4%
Long-term bonds and mortgages	26.5	64.7	23.5	59.5
Common stocks	10.0	24.3	11.5	29.1
	$41.0	100.0%	$39.5	100.0%

Investment Performance		
	Annual Rates of Return for Periods Ending 9/30/2028	
	5 Years	1 Year
Total Alpine Fund:		
Time-weighted	8.2%	5.2%
Dollar-weighted (internal)	7.7%	4.8%
Assumed actuarial return	6.0%	6.0%
U.S. Treasury bills	7.5%	11.3%

	Annual Rates of Return for Periods Ending 9/30/2028	
	5 Years	1 Year
Large sample of pension funds (average 60% equities, 40% fixed income)	10.1%	14.3%
Common stocks—Alpine Fund	13.3%	14.3%
Alpine portfolio beta coefficient	0.90	0.89
Standard & Poor's 500 stock index	13.8%	21.1%
Fixed-income securities—Alpine Fund	6.7%	1.0%
Broad Investment Grade bond index	4.0%	−11.4%

Karl was proud of his performance and was chagrined when a trustee made the following critical observations:

a. "Our 1-year results were terrible, and it's what you've done for us lately that counts most."

b. "Our total fund performance was clearly inferior compared to the large sample of other pension funds for the last five years. What else could this reflect except poor management judgment?"

c. "Our common stock performance was especially poor for the 5-year period."

d. "Why bother to compare your returns to the return from Treasury bills and the actuarial assumption rate? What your competition could have earned for us or how we would have fared if invested in a passive index (which doesn't charge a fee) are the only relevant measures of performance."

e. "Who cares about time-weighted return? If it can't pay pensions, what good is it?"

Appraise the merits of each of these statements and give counterarguments that Mr. Karl can use.

3. The Retired Fund is an open-ended mutual fund composed of $500 million in U.S. bonds and U.S. Treasury bills. This fund has had a portfolio duration (including T-bills) of between 3 and 9 years. Retired has shown first-quartile performance over the past five years, as measured by an independent fixed-income measurement service. However, the directors of the fund would like to measure the market timing skill of the fund's sole bond investor manager. An external consulting firm has suggested the following three methods:

a. Method I examines the value of the bond portfolio at the beginning of every year, then calculates the return that would have been achieved had that same portfolio been held throughout the year. This return would then be compared with the return actually obtained by the fund.

b. Method II calculates the average weighting of the portfolio in bonds and T-bills for each year. Instead of using the actual bond portfolio, the return on a long-bond market index and T-bill index would be used. For example, if the portfolio on average was 65% in bonds and 35% in T-bills, the annual return on a portfolio invested 65% in a long-bond index and 35% in T-bills would be calculated. This return is compared with the annual return that would have been generated using the indexes and the manager's actual bond/T-bill weighting for each quarter of the year.

c. Method III examines the net bond purchase activity (market value of purchases less sales) for each quarter of the year. If net purchases were positive (negative) in any quarter, the performance of the bonds would be evaluated until the net purchase activity became negative (positive). Positive (negative) net purchases would be viewed as a bullish (bearish) view taken by the manager. The correctness of this view would be measured.

Critique *each* method with regard to market timing measurement problems.

Use the following data to solve CFA Problems 4 and 5: The administrator of a large pension fund wants to evaluate the performance of four portfolio managers. Each portfolio manager invests only in U.S. common stocks. Assume that during the most recent 5-year period, the average annual total rate of return including dividends on the S&P 500 was 14%, and the average nominal rate of return on government Treasury

bills was 8%. The following table shows risk and return measures for each portfolio:

Portfolio	Average Annual Rate of Return	Standard Deviation	Beta
P	17%	20%	1.1
O	24	18	2.1
R	11	10	0.5
S	16	14	1.5
S&P 500	14	12	1.0

4. What is the Treynor performance measure for portfolio *P*?

5. What is the Sharpe performance measure for portfolio *Q*?

6. An analyst wants to evaluate portfolio *X*, consisting entirely of U.S. common stocks, using both the Treynor and Sharpe measures of portfolio performance. The following table provides the average annual rate of return for portfolio *X*, the market portfolio (as measured by the S&P 500), and U.S. Treasury bills during the past 8 years:

	Average Annual Rate of Return	Standard Deviation of Return	Beta
Portfolio *X*	10%	18%	0.60
S&P 500	12	13	1.00
T-bills	6	N/A	N/A

   a. Calculate the Treynor and Sharpe measures for both portfolio *X* and the S&P 500. Briefly explain whether portfolio *X* underperformed, equaled, or outperformed the S&P 500 on a risk-adjusted basis using both the Treynor measure and the Sharpe ratio.

   b. On the basis of the performance of portfolio *X* relative to the S&P 500 calculated in part (*a*), briefly explain the reason for the conflicting results when using the Treynor measure versus the Sharpe ratio.

7. Assume you invested in an asset for two years. The first year you earned a 15% return, and the second year you earned a negative 10% return. What was your annual geometric return?

8. A portfolio of stocks generates a −9% return in 2018, a 23% return in 2019, and a 17% return in 2020. What was the annualized return (geometric mean) for the entire period?

9. A 2-year investment of $2,000 results in a cash flow of $150 at the end of the first year and another cash flow of $150 at the end of the second year, in addition to the return of the original investment. What is the dollar-weighted (internal) rate of return on the investment?

10. In measuring the performance of a portfolio, the time-weighted rate of return may be preferred to the dollar-weighted rate of return because:

    a. When the rate of return varies, the time-weighted return is higher.

    b. The dollar-weighted return assumes all portfolio deposits are made on day 1.

    c. The dollar-weighted return can only be estimated.

    d. The time-weighted return is unaffected by the timing of portfolio contributions and withdrawals.

11. A pension fund portfolio begins with $500,000 and earns 15% the first year and 10% the second year. At the beginning of the second year, the sponsor contributes another $500,000. What were the time-weighted and dollar-weighted rates of return?

12. During the annual review of Acme's pension plan, several trustees questioned their investment consultant about various aspects of performance measurement and risk assessment.

    a. Comment on the appropriateness of using each of the following benchmarks for performance evaluation:
    - Market index.
    - Benchmark normal portfolio.
    - Median of the manager universe.

    b. Distinguish among the following performance measures:
    - The Sharpe ratio.
    - The Treynor measure.
    - Jensen's alpha.

      i. Describe how each of the three performance measures is calculated.

      ii. State whether each measure assumes that the relevant risk is systematic, unsystematic, or total. Explain how each measure relates excess return and the relevant risk.

13. Trustees of the Pallor Corp, pension plan ask consultant Donald Millip to comment on the following statements. What should his response be?

    a. Median manager benchmarks are statistically unbiased measures of performance over long periods of time.

    b. Median manager benchmarks are unambiguous and are therefore easily replicated by managers wishing to adopt a passive/indexed approach.

    c. Median manager benchmarks are not appropriate in all circumstances because the median manager universe encompasses non-uniform investment styles.

14. James Chan is reviewing the performance of the global equity managers of the Jarvis University endowment fund. Williamson Capital is currently the endowment fund's only large-capitalization global equity manager. Performance data for Williamson Capital are shown in Table 8A.

Chan also presents the endowment fund's investment committee with performance information for Joyner Asset Management, which is another large-capitalization global equity manager. Performance data for Joyner Asset Management are shown in Table 8B. Performance data for the relevant risk-free asset and market index are shown in Table 8C.

Table 8A **Williamson Capital Performance Data, 2020–2025**

Average annual rate of return	22.1%
Beta	1.2
Standard deviation of returns	16.8%

Table 8B **Joyner Asset Management Performance Data, 2020–2025**

Average annual rate of return	24.2%
Beta	0.8
Standard deviation of returns	20.2%

Table 8C **Risk-Free Asset and Market Index Performance Data, 2020–2025**

Risk-Free Asset	
Average annual rate of return	5.0%
Market Index	
Average annual rate of return	18.9%
Standard deviation of returns	13.8%

a. Calculate the Sharpe ratio and Treynor measure for both Williamson Capital and Joyner Asset Management.

b. The investment committee notices that using the Sharpe ratio versus the Treynor measure produces different performance rankings of Williamson and Joyner. Explain why these criteria may result in different rankings.

E-INVESTMENTS EXERCISES

Morningstar has an extensive ranking system for mutual funds, including a screening program that allows you to select funds based on a number of factors. Open the Morningstar Web site at **www.morningstar.com** and click on the *Mutual Funds* link. Select the *Fund Quickrank* link in the Performance section. (Free registration is required to access the site.) Use the Quickrank screener to find a list of large growth stock funds with the highest 1-year returns. Repeat the process to find the funds with the highest 3-year returns. What fraction of funds appear on both lists?

Select three of the funds that appear on both lists. For each fund, click on the ticker symbol to get its Morningstar report and look in the Ratings & Risk section.

1. What is the fund's standard deviation?
2. What is the fund's Sharpe ratio?
3. What is the fund's Treynor ratio?
4. What is the standard index? What is the best-fit index?
5. What are the beta and alpha coefficients using both the standard index and the best-fit index? How do these compare to the fund's parameters?
6. Look at the Management section of the report. Was the same manager in place for the entire 10-year period?
7. Are any of these funds of interest to you? How might your screening choices differ if you were choosing funds for various clients?

The following questions are intended to help candidates understand the material. They are *not* actual FRM exam questions.

SOLUTIONS TO CONCEPT CHECKS

8.1.

Time	Action	Cash Flow
0	Buy two shares	−40
1	Collect dividends; then sell one of the shares	4 + 22
2	Collect dividend on remaining share, then sell it	2 + 19

a. Dollar-weighted return:

$$-40 + \frac{26}{1 + r} + \frac{21}{(1 + r)^2} = 0$$

$$r = .1191, \text{ or } 11.91\%$$

b. Time-weighted return:

The rates of return on the stock in the 2 years were:

$$r_1 = \frac{2 + (22 - 20)}{20} = .20$$

$$r_2 = \frac{2 + (19 - 22)}{22} = -.0455$$

Arithmetic time-weighted return:

$(r_1 + r_2)/2 = 0.773$, or 7.73%

Geometric time-weighted return:

$[(1 + r_1)(1 + r_2)]^{1/2} - 1 = 0.702$

$= 7.02\%$

8.2. Sharpe: $(\bar{r} - r_f)/\sigma$

$S_P = (35 - 6)/42 = .69$

$S_M = (28 - 6)/30 = .733$

Alpha: $\bar{r} - [\bar{r}_f + \beta(\bar{r}_M - r_f)]$

$\alpha_P = 35 - [6 + 1.2(28 - 6)] = 2.6$

$\alpha_M = 0$

Treynor: $(\bar{r} - \bar{r}_f)/\beta$

$T_P = (35 - 6)/1.2 = 24.2$

$T_M = (28 - 6)/1.0 = 22$

Information ratio: $\alpha/\sigma(e)$

$I_P = 2.6/18 = .144$

$I_M = 0$

Therefore, portfolio *P* outperformed the market according to the Jensen, Treynor, and information measures, but had an inferior Sharpe measure.

8.3. The alpha exceeds zero by .2/2 = .1 standard deviations. A table of the normal distribution (or, somewhat more appropriately, the distribution of the *t*-statistic) indicates that the probability of such an event, if the analyst actually has no skill, is approximately 46%.

8.4. The timer will guess bear or bull markets completely randomly. One-half of all bull markets will be preceded by a correct forecast, and similarly for bear markets. Hence $P_1 + P_2 - 1 = ½ + ½ - 1 = 0$.

8.5. First compute the new bogey performance as (.70 × 5.81) + (.25 × 1.45) + (.05 × .48) = 4.45.

a. Contribution of asset allocation to performance:

	(1)	(2)	(3)	(4)	(5) = (3) × (4)
Market	**Actual Weight in Market**	**Benchmark Weight in Market**	**Active or Excess Weight**	**Market Return (%)**	**Contribution to Performance (%)**
Equity	0.70	0.70	0.00	5.81	0.00
Fixed-income	0.07	0.25	−0.18	1.45	−0.26
Cash	0.23	0.05	0.18	0.48	0.09
Contribution of asset allocation					−0.17

b. Contribution of selection to total performance:

	(1)	(2)	(3)	(4)	(5) = (3) × (4)
Market	**Portfolio Performance (%)**	**Index Performance (%)**	**Excess Performance (%)**	**Portfolio Weight**	**Contribution (%)**
Equity	7.28	5.00	2.28	0.70	1.60
Fixed-income	1.89	1.45	0.44	0.07	0.03
Contribution of selection within markets					1.63

9

Hedge Funds

Learning Objectives

After completing this reading you should be able to:

- Explain biases that are commonly found in databases of hedge funds.
- Explain the evolution of the hedge fund industry and describe landmark events that precipitated major changes in the development of the industry.
- Explain the impact of institutional investors on the hedge fund industry and assess reasons for the growing concentration of assets under management (AUM) in the industry.
- Explain the relationship between risk and alpha in hedge funds.
- Compare and contrast the different hedge fund strategies, describe their return characteristics, and describe the inherent risks of each strategy.
- Describe the historical portfolio construction and performance trends of hedge funds compared to those of equity indices.
- Describe market events that resulted in a convergence of risk factors for different hedge fund strategies and explain the impact of such convergences on portfolio diversification strategies.
- Describe the problem of risk sharing asymmetry between principals and agents in the hedge fund industry.

Excerpt is reproduced from William Fung and David A. Hsieh, Handbook of the Economics of Finance, *Volume 2B: Financial Markets and Asset Pricing, 9780444535948, 2013, Constantinides et al. Ch 16, Pg 1063-1125.*

9.1 THE HEDGE FUND BUSINESS MODEL—A HISTORICAL PERSPECTIVE

The success of an investment vehicle depends on investors' perception of its performance, how the investment manager operates the portfolio and the efficiency of the vehicle's organization structure. A publicly traded, indexed fund, for example, has well defined performance targets, often managed to have holdings that mimic a benchmark index and structured in a format to comply with the listing regulations. Hedge funds, on the other hand, have their roots in the world of private wealth management. For over half a century[1], wealthy individuals invested their capital alongside "talented" traders expecting out-sized return to their investment irrespective of general market conditions.[2]

Hedge funds are distinct from mutual funds in several important respects. Historically, hedge funds are private investment vehicles not open to the general investment public. This means that hedge funds face less regulation than publicly traded mutual funds, allowing them to hold substantial short positions to preserve capital during market downturns.[3] Typically hedge fund managers generate profit from both long as well as short positions. However, some specialist hedge fund managers, who are particularly skilled in identifying "over-priced" assets and have the infrastructure to carry short positions over an extended period of time, do rely on shorting securities as their main source of profit.[4] The ability to take short positions not only helps to dampen sensitivity to the general market direction, it also allows managers to take large bets on perceived relative price discrepancies of assets. Therefore, it is common to find hedge fund balance sheets that substantially exceed in size the equity capital of the vehicles.[5] It is the practice of shorting and the leveraging of investors' capital that distinguish hedge funds from conventional long-bias funds. Managing leveraged positions in volatile markets to preserve capital calls for skillful management of position size. To deliver persistent positive returns requires market-timing skills. These are the defining characteristics of a hedge fund manager's skill set—the ability to identify profitable long as well as short opportunities in a range of asset categories, the organization structure to carry short positions for extended periods of time,[6] the know-how to fund leveraged positions, and the risk management skill to maintain complex positions during volatile markets.[7]

The private nature of hedge funds often suits both the needs of investors and managers. While wealthy investors of early hedge fund vehicles rarely impose specific mandates on how their investments should be managed, most if not all of these investors will demand their investment in the vehicle be kept private and carry limited liability. In addition, as individual capital commitments to a hedge fund manager tend to be small in relation to the investor's overall portfolio, it is critical that a hedge fund investment carries the same limited liability protection as buying shares of a public company. This is especially important given the leverage used by most hedge fund managers.

Hedge fund managers often claim to have complex proprietary strategies to generate outsized profits. To keep other traders from mimicking or "front running" their trades, they offer very little transparency, even to their investors. The opacity of hedge fund vehicles persisted for over half a century until the arrival of institutional investors in the new millennium.

Another benefit of being lightly regulated investment vehicles is that hedge funds in the US are not subject to the legal restrictions on compensation that govern publicly traded mutual funds. A typical hedge fund charges a fixed management fee, which usually ranges between one and two percent per annum, based on the value of assets they manage. The lower end of this range is comparable to the management fees charged by actively managed mutual funds. However, unlike mutual funds, hedge funds generally charge an incentive fee—typically between 10 and 20% of new profits of the fund. Incentive fees are only payable when new profits are made. This means that losses have to be carried forward until they are recouped and the previous level of investors capital is restored—this level of capital is often referred to as the High Water Mark.

However, being lightly regulated does have unintended consequences. For one, performance records of hedge funds are generally not standardized and available reports are prone to

[1] It is generally accepted that Alfred W. Jones started the first hedge fund in 1949.

[2] We believe this co-investing mentality had a profound influence on the contractual relationship between hedge fund managers and their investors, a great deal of which survived to the modern day.

[3] Although there is no generally accepted explanation to why these pools of capital came to be known as "hedge funds," one plausible explanation may be the need to go short so as to meet investors' expectation of absolute profit from their investments closely resembles the concept of hedging.

[4] A group of managers often referred to as *Short Sellers*.

[5] This is analogous to investment banks whose trading positions are often leveraged.

[6] See Fung and Hsieh (2011).

[7] Prior to the 2008 financial market crisis, this was an often underrated skill of hedge fund managers despite some famous examples in which otherwise profitable positions have to be liquidated prematurely leading to a complete failure of the fund.

measurement errors.[8] Until the early 1990s, the historical performance of hedge funds was often as private as their investors, and the assessment of hedge fund performance was as much art as science. By 1993, reports of hedge fund managers amassing billions of dollars of leverageable capital began to emerge.[9] Unfortunately this industry landmark was followed by dramatic losses from some well-known managers in the hedge fund industry in early 1994, triggered by the unexpected change in interest rate policy by the Federal Reserve.[10] The events of 1994 had a profound impact on the way hedge fund investors assess their portfolios. It was the first recorded event in which the proverbial alarm—*everyone lost money*—rang throughout the hedge fund industry. Coupled with the growth of the hedge fund industry, the events of 1994 prompted investors to re-examine the perceived diversification of their hedge fund holdings. In turn, this aided the development of electronically available hedge fund databases with which more formal analysis of hedge fund performance and the attendant risks can be conducted.[11] It was the arrival of electronic databases that made academic research in hedge funds a feasible proposition. This is also where our story begins.

Just how big is the hedge fund industry? We start with the commercial hedge fund databases. Hedge funds are not required to disclose information to the general public. While some hedge fund firms consistently report data on their funds to multiple databases, many report only selected samples of their funds to one database, and an unknown number do not report to any database, past and present. This lack of a performance reporting standard creates some unique challenges for researchers. In order to gain insight on the capital formation of the hedge fund industry, we need the broadest coverage of the industry possible. Often researchers have to merge multiple databases. Our hedge fund data source in this chapter combines three commonly used commercial hedge fund databases—BarclayHedge, HFR, and Lipper-Tass. Table 9.1 reports the number of funds, as well as the assets under management (AUM).[12] The table shows that the number of funds increased more than fourfold from 1997 to 2010, while the AUM went up nearly tenfold.

While there is no systematic way to ascertain the number of hedge funds that do not participate in commercial databases, there are independent industry surveys that provide helpful clues. In particular HFN (Hedgefund.net) has been conducting surveys on a comprehensive sample of administrators of hedge fund vehicles since 2003.[13] From the survey data, we can see that the AUM of the funds in the commercial databases is approximately 50–60% of the hedge fund assets serviced by administrators (or assets under administration, "AUA".

The rapid growth of industry AUM has been accompanied by a high turnover of hedge funds and their managers. We can get a glimpse of this from the databases. For the sample period 1997 until 2010, the average annual entry rate for hedge funds into the commercial databases is 9% whereas the "exit" rate is much higher. On average, funds in the databases stop reporting at the rate of 21% per year. However, exiting a database does not necessarily mean the termination or liquidation of a hedge fund. Since hedge funds voluntarily report to commercial databases,

[8] Unlike mutual funds which have had to conform to a reporting standard and well-defined disclosure rules for many years, hedge funds have escaped the purview of financial regulators for decades. Two major events are gradually changing the operating standard of hedge funds. First, regulators around the world are beginning to issue guidance on "best practice" for private investment vehicles and custodian/administrative service providers to the hedge fund community have consolidated into a small number of large companies. An "acceptable operating standard" is beginning to emerge among hedge funds. Industry organizations such as the Chartered Alternative Investment Analysts (CAIA) Association also help to define "best practice" guidelines for professionals operating in the hedge fund industry—see http://www.caia.org. Nonetheless, in analyzing historical performance, it is important to be aware of return measurement biases. See for example, Bollen and Pool (2009), Cassar and Gerakos (2011), and Getmansky, Lo, and Makarov (2004).

[9] Both HFR and Lipper-Tass regularly release statistical reports on the hedge fund industry—they are the HFR Global Hedge Fund Industry Report (see https://www.hedgefundresearch.com/index.php?fuse=products-irmm&1301697978) and the Lipper-Tass Asset Flows Report. Standard statistics on capital flows (contribution and redemption data net of performance) and industry size are tabulated in these reports. Different industry sources do not always agree on industry statistics. For example, by 1993 HFR estimated the total assets managed by the hedge fund industry to be just shy of $100 billion, whereas Lipper-Tass estimated the size of the industry to be just below $50 billion. In general, historical industry size statistics are notoriously noisy; mostly a consequence of the voluntary nature of hedge fund reporting.

[10] See Fung and Hsieh (2000b).

[11] It was around this time that the three oldest hedge fund databases began their collection of hedge fund performance records—these are CISDM (Center for International Securities and Derivative Markets, formerly Mar/Hedge), HFR and Lipper-Tass. Although there were other private databases such as the Offshore Fund Directory (used in Brown, Goetzmann, and Ibbotson (1999), the aforementioned three remain the earliest surviving entrants to the arena of electronic databases that are commercially available to the public.

[12] The Barclayhedge database can be found at www.barclayhedge.com. As much as possible, duplicate funds are eliminated to avoid double counting.

[13] As the location of funds such as offshore funds are frequently different from the location of the hedge fund manager's operations, it is common practice for hedge funds to engage the services of professional firms which provide local corporate services such as company directors, book keeping and independent valuation of the funds' assets. Collectively these services are often provided by independent administrative service providers who assume these corporate duties for the hedge fund vehicle. This is distinct from the role of the hedge fund manager who acts as the investment advisor to the hedge fund vehicle that houses investors capital.

Table 9.1 Number of Unique Funds and AUM After Merging Three Commercial Databases (BarclayHedge, HFR, and Lipper-Tass), and Assets Under Administration (AUA)

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
# of funds:	1,205	1,276	1,138	1,312	1,970	2,268	2,642	3,248	3,907	4,310	5,108	5,309	4,858	6,182
AUM ($b):	156	161	198	217	202	262	343	483	622	779	1107	772	769	945
HFN	—	—	—	—	—	—	845	1,293	1,531	2,153	2,861	1,932	2,172	2,826
AUA ($b)														

Source: BarclayHedge, HFR, Lipper-Tass, Hedgefund.net.

they can also choose to stop reporting for a variety of reasons. When a fund stops reporting to a commercial database, whenever possible the database vendor provides an explanation of this event. In about one third of these cases, a fund stops reporting because it has lost so much assets that it is no longer a viable business and the fund is liquidated by the manager. However, in the other two thirds of cases, the explanation of the exit is incomplete at best. For example, it is entirely possible that an operating fund may elect to stop reporting to databases because it has attracted sufficient capital and the hedge fund manager no longer seeks new investors. Disclosing the fund's statistics becomes a costly exercise with little benefit. This is, of course, diametrically opposed to cases in which things are going so badly in a fund that reporting to databases is not a priority in the struggle for survival—many of which may eventually cease operations, but some could survive. Add to the fact that there are those hedge fund firms that never disclosed their information to commercially available databases, it is far from obvious that what we learn about hedge funds in these databases applies to the entire population. This is commonly referred to as *selection bias* or *self-reporting bias* in hedge fund databases.

However, we do have one small piece of evidence suggesting that the selection bias in commercial databases may not be very large. This evidence is based on the average return of funds-of-hedge funds (FOHF for short). FOHFs are vehicles that invest in a portfolio of individual hedge funds offering investors a one-stop diversification service to the hedge fund industry for a fee. FOHFs do not limit their investment to those hedge funds found in commercial databases. Assuming that FOHFs invest in the entire population of hedge funds, including those that do not participate in commercial databases, their average return should better reflect the average return of the entire population of hedge funds. This alternative approach to assessing the hedge fund industry as a whole was proposed in Fung and Hsieh (2000a). The fact that the average return of FOHFs is highly correlated to the average return of hedge funds in commercial databases suggests that the latter is not a special subset of the entire population of hedge funds. However, until we can systematically and significantly increase the coverage of assets invested in the hedge fund industry beyond commercially available databases, empirical conclusions derived from these databases will continue to be overshadowed by the missing assets managed by nonreporting hedge fund managers.[14]

9.2 EMPIRICAL EVIDENCE OF HEDGE FUND PERFORMANCE

The first few academic studies of hedge funds came soon after hedge fund databases were made publicly available—they are Fung and Hsieh (1997a, 1997b), Ackermann, McEnally, and

[14] Before proceeding, we offer some comments regarding potential ways to extend hedge fund coverage beyond commercial databases. Let us use the term "reporting hedge funds" to denote those funds that participate in at least one commercial database, and "non-reporting hedge funds" to denote those that do not participate in any commercial database. Some FOHFs are registered and regulated vehicles. They are required to disclose publicly some information on the hedge funds in their portfolios usually on a quarterly interval. These disclosures typically contain the names and pricing of their investments in reporting as well as nonreporting hedge funds. Unfortunately, it is not obvious how to correctly convert the pricing information to return information based on quarterly snapshots of portfolios, since interim trades are not reported. Even if one manages to create quarterly returns, they are only available for the duration of the filings. Information on strategy, fees, performance and AUM histories would be incomplete or missing. Another avenue for finding non-reporting hedge funds is to comb through Form ADV filings of large money management firms. Here, the researcher would encounter the problem that Form ADV does not require a filer to designate whether a fund on its form is a hedge fund or some other type of private investment such as private equity or venture capital. Even if one can identify that a filer's vehicle is a hedge fund, one still has the daunting task of locating strategy, fees, performance and AUM histories of non-reporting hedge funds. Lastly, we note that management firms that operate hedge funds are not always required to file Form ADV—only those with more than 14 US investors, having assets of at least US$ 25 million, and having a lockup period less than two years. Thus, Form ADV cannot be viewed as a "census" of hedge funds in the global hedge fund industry. In any case, even if the names of non-reporting hedge funds can be found in Form ADV, one still has to locate their performance and AUM histories from some other source, as they are not in commercial databases.

00011693-00000002_CH09_p147-186.indd 150 05/07/22 3:06 PM

Ravencraft (1999) and Brown, Goetzmann, and Ibbotson (1999). Although these early studies vary in emphasis, all four share a common concern regarding potential measurement errors and the attendant biases in reported hedge fund returns. All of these biases emanate from the lack of a performance reporting standard among hedge funds and the self-selection nature of reporting data to database vendors.[15] These return measurement biases impact different performance metrics in different ways and the analysis of these biases is an integral part of how we assess historical hedge fund returns.

Were the Lofty Expectations of Early Hedge Fund Investors Fulfilled?

In a recent review of hedge funds, Stulz (2007) pointed to the out-performance of the then Credit Suisse/Tremont Hedge Fund index ("CTI" for short) compared to two major equity indices—the Standard & Poor's 500 ("SNP" for short) and the Financial Times World Index—in both cumulative return as well as standard deviation of returns over the period 1994 to the middle of 2006.[16] While this is a supportive observation for the rise in hedge fund popularity among equity-oriented investors, more needs to be said about the factors that propelled the growth of the hedge fund industry prior to 1994, incorporating what we have learned about measurement biases in hedge fund indices. Specifically, how much of the previously reported favorable hedge fund performance is due to measurement biases and how much can be explained by risk differences? Put differently, adjusting for measurement biases and risk, do hedge funds on average deliver alpha and what should be the reference risk factor(s)?[17] We begin with a few observations on the history of indices of hedge fund performance.

Broad-based indices of hedge fund performance became available with the arrival of commercial hedge fund databases. For instance, Hedge Fund Research ("HFR" for short) publishes a range of hedge fund indices designed to provide an average of the industry performance as well as sub-indices that reflect specialized strategies and sectors. Most of the HFR indices have histories dating back to 1990. However, HFR itself only started collecting data from hedge fund managers sometime around 1994. Therefore, pre-1994 data is mostly "backfilled".[18] Coupled with the fact that reporting performance to databases is voluntary, the risk exists that observable data are biased in favor of good performance.[19] This effect was partially mitigated when the collection of performance data went live (circa 1994).[20] Consequently one may be justified in being skeptical about hedge fund indices good performance prior to 1994. In addition, during the early days of collecting hedge fund data, it is quite likely that the scope of manager coverage is limited relative to the later years. Although when funds begin reporting to databases, its prior returns will become available and can be "incorporated" into an index after-the fact, this will expose the index returns to backfill bias.[21] Taken together, the empirical evidence suggests that hedge fund index returns became more reliable around 1996.

But then what about the performance prior to 1996? To answer this question, we appeal to earlier studies that tracked the performance of well-known large hedge funds. In two separate studies, Fung and Hsieh (2000b) and Fung, Hsieh, and Tsatsaronis (2000) collected performance data of 27 hedge funds, each with more than US$1 billion of assets at the end of 1997, to measure the market impact of their trading. Note that seven out of these 27 large hedge funds, just over 25%, did not report to commercial databases. Their data had to be hand collected from investors' sources; see Table 9.1 of Fung and Hsieh (2000b).

Figure 9.1 charts the cumulative performance of these 27 large hedge funds compared to the SNP index from 1987 until 1996. It shows that these large hedge funds easily outperformed the

[15] For a summary of these measurement biases and their impact on performance metrics see Fung and Hsieh (2000a).

[16] See p. 183 of Stulz (2007).

[17] We are not suggesting that alpha is the final quest of hedge fund investors, rather it is that dimension of performance that distinguishes hedge fund returns from conventional risk premium.

[18] In the sense that historical data are collected after-the-fact the accuracy of which critically depend on the willingness and ability of hedge fund managers to provide data.

[19] For example if a manager operates several hedge funds, it is questionable whether the poor performing ones will find their way into databases. In other words, there may well be a tendency to "put the best face forward". The limit of this type of self-selection bias is the well-known survivorship bias. Simply put, poor performing funds that have been terminated cannot be readily observed.

[20] But only to the extent that the inclusion and exclusion criteria of a database themselves do not lead to omissions of relevant observations. For example some databases may question the accuracy of a fund's reported returns due to the tardiness of reporting frequency and choose to drop the fund in question from the database. This type of behavior is common among funds that find themselves in difficulties and prefer not to continue to release "bad news" to the public. Therefore the observed returns from a database may only capture part of the downward spiral of a dying fund.

[21] A related and important form of backfill bias is sometime refered to as the "*instant history*" bias. In essence when a new fund enters the database some of its performance history during its incubation period is incorporated without clear distinction from the live performance data going forward. Since unsuccessful incubations will not present themselves to databases, this potentially creates an upward bias to the historical returns of new funds. For a more detailed discussion of this issue see Fung and Hsieh (2009), and Aggarwal and Jorion (2010).

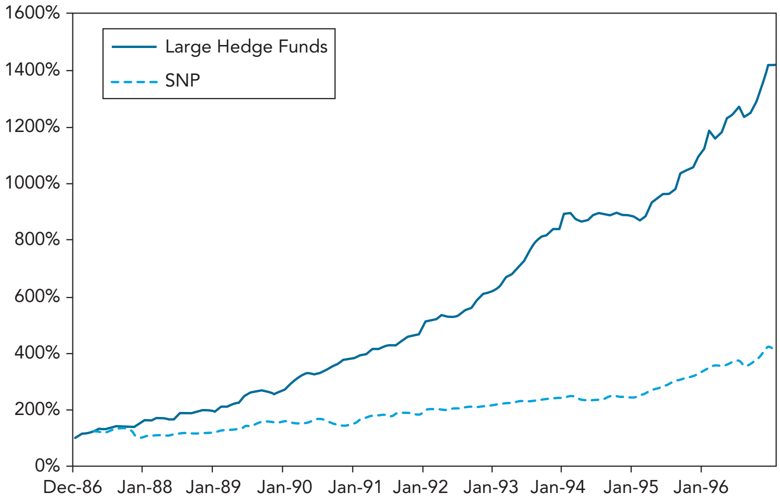


Figure 9.1 **Cumulative returns of large hedge funds and the SNP index: 1987–1996.**

Source: Fung and Hsieh, 2000b; Fung, Hsieh, and Tsatsaronis, 2000.

SNP index by a wide margin. However, this outperformance has to be considered with caution. Clearly, we are able to identify these 27 large hedge funds because they succeeded and became visible at the end of the sample period. Therefore, it is not surprising that they exhibit outstanding performance. This is an early example of what is now widely recognized in the hedge fund industry as survivorship (that these funds did not liquidate) and selection (that their AUMs are based on end of the period data) biases. The magnitude of these biases during the early days of the hedge fund industry remains unknown to us. Empirical estimates of these biases only surfaced after the arrival of commercial databases whose early data is exposed to the same measurement problems. Nonetheless, these performance data are informative and insight can be gleaned from them.

First, during the early days of the hedge fund industry, a substantial amount of the industry's assets was controlled by a small number of large hedge funds.[22] Second, before the arrival of commercial databases, selective reports in the financial press tend to gravitate towards more newsworthy large hedge funds. It is conceivable that these statistics may figure prominently in investors' perception of the hedge fund industry's performance. Third, despite the fact that investors' perception of the return from these lightly regulated, opaque investment vehicles may well have been seen through rosetinted lenses, the spectacular results in Figure 9.1 show a very sizeable performance gap between the large hedge funds and the SNP, which cannot be easily accounted for by just measurement biases. It is a reasonable conjecture that early investors of the hedge fund industry were attracted to potential out-sized returns during the post-87 to pre-94 period. Another interesting point to note is that there may well be a sizeable performance gap between successful large funds and those that struggled to growth assets. We will return to this point when we discuss the capital formation trend of the hedge fund industry. For now, suffice to say that Figure 9.1 shows that these early investors were amply rewarded despite major market events such as the Savings and Loans crisis and the first Gulf War.

Post the events of 1994, it became clear to investors that hedge fund managers operating seemingly different strategies can end up with similar market exposures which can go sour at the same time—we call this hedge fund convergence risk.[23] When this occurs, the diversification of a hedge fund portfolio implodes, and portfolio risk converges to a small number of highly concentrated systemic risk factors. Consequently a portfolio strategy of betting on a handful of large hedge funds relying on the individual managers to diversify away systemic risks can be a very risky proposition. In the ensuing years investors became more active in broadening their portfolio diversification to control both dimensions of risks—manager specific as well as market factors.[24] Intermediaries such as funds-of-hedge-funds, whose purpose is to offer hedge fund investors a diversified portfolio of hedge funds in a single investment vehicle albeit for an additional fee, emerged and flourished. Competing for investors' capital, the arrival of intermediaries led to better disclosure and improved transparency in the hedge fund industry. Together with the arrival of hedge fund indices, performance reporting began to standardize and peer-group averages of hedge funds performance became more readily available. To date, there are two major indices with the longest real-time reporting history: they are the Dow Jones-Credit Suisse Broad Index (formerly the

[22] Table 1 of Fung and Hsieh (2000b) estimates the total AUM by these 27 funds at $63.94 billion. Early estimates of the hedge fund industry's size range from $89.9 billion to $146 billion; see Eichengreen et al. (1998). Therefore these 27 funds' AUM alone represent a significant percentage of the total industry AUM.

[23] More discussion on how different hedge fund strategies can lead to similar factor bets can be found in a later section.

[24] More discussions on how this can be achieved can be found in the later section on portfolio construction.

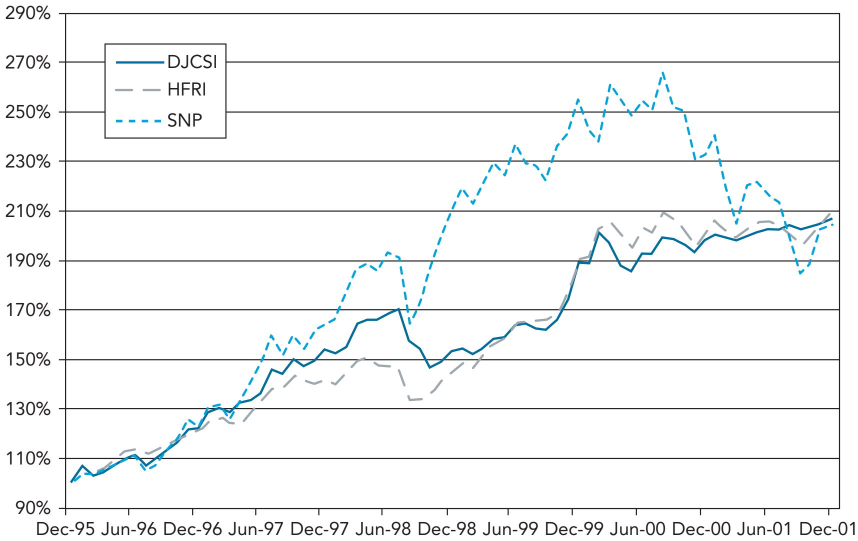


Figure 9.2 **Cumulative return of DJCSI, HFRI, and the SNP index: 1996–2001.**

Credit Suisse/Tremont Index) and the Hedge Fund Research Fund Weighted Composite Index (respectively "DJ CSI" and "HFRI" for short).[25] Figure 9.2 charts the cumulative performance of these two indices and the SNP index.

As discussed earlier, pre-1996 the data used to compute publicly available hedge fund indices are susceptible to measurement biases. Consequently we track the historical performance of hedge fund indices using data from 1996 onwards. We do this by analyzing sub-period performance separated by market events identified in Fung et. al. (2008). Moving forward in time, the first major market event is the collapse of the well-known hedge fund Long Term Capital Management ("LTCM") in the fall of 1998. Accordingly we begin by examining performance from January 1996 to July 1998 just before the stressful months leading up to the demise of LTCM. Figure 9.2 shows that over the period January 1996 to July 1998, a broadly diversified portfolio of hedge funds did not outperform the SNP index—which is in sharp contrast to the results in Figure 9.1. Cumulative returns over this period are 170.39%, 146.31%, and 190.97% respectively for the DJCSI, HFRI and SNP indices. However, a case can be made in favor of hedge fund investing in terms of risk characteristics. The annualized standard deviations are 10.87%, 8.90%, and 16.21% respectively for the DJCSI, HFRI, and SNP index.

The collapse of a large, well-known hedge fund like LTCM had a dramatic impact on the private world of hedge fund investors.[26] It is a stark reminder that earning outsized returns from highly leveraged bets comes with spectacular event risks—namely, outsized draw downs that wiped out LTCM investors' capital. This was an extreme event that did not have the same adverse effect on the equity market—during the worst months of the LTCM crisis from June to October of 1998, the DJCSI and HFRI lost 13.04% and 7.68% respectively whereas the SNP index only lost 2.62%. The circumstances leading up to the collapse of LTCM have been referred to "a ten sigma event," which may well be true, nonetheless the dramatic loss from a well-known fund on such a major scale left an indelible scar on hedge fund investors' confidence in hedge funds. We believe this was a turning point in the capital formation process of the hedge fund industry.

According to the Lipper-Tass[27] Asset Flow Report, net of performance differences, investors increased their investments in the hedge fund industry by 45.62% over the 1996–1997 period. In contrast, the comparable figure for the 1998–1999 period, despite the dramatic performance recovery in 1999,[28] dropped to 8.48%; reflecting a significant reduction in investors' appetite for hedge funds. This LTCM event raised many questions on how the risk of hedge fund investments should be measured. Many ad hoc statistics have been proposed to measure the tail risk of investing in highly leveraged hedge fund strategies, a subject that we will return to later after a more thorough analysis of the inherent risk factors of different hedge fund strategies. Suffice to say that by the end of 1998, the hedge fund

[25] The DJCSI was previously known as the Credit Suisse/Tremont Index (CTI) and is a weighted average of component hedge fund's performance, see www.hedgeindex.com for details on the index construction plus return history and www.hfr.com for the HFRI index construction plus return history.

[26] Most of the well-known private Swiss banks who collectively handle substantial amounts of the world's high-net-worth investors' wealth were major LTCM's investors; see Lowenstein (2001).

[27] The original TASS database was sold to Tremont Asset Management which subsequently became the Credit Suisse/Tremont database. The database part of the business was sold to Lipper Financial Services (a wholly owned company of Thomson Reuters) while the index production business was retained by Credit Suisse and is now a joint venture between Credit Suisse and Dow Jones Indices (a CME group company).

[28] For the calendar year 1999, DJCSI and HFRI returned 21.65% and 27.89% respectively which compare favorably to the SNP index's return of 20.03%.

investment community had generally accepted that the first two moments of an expected return distribution may be woefully inadequate for capturing the risk of these dynamic, nonlinear, leveraged strategies.

The Arrival of Institutional Investors

The ensuing two year period, 2000–2001, witnessed the burst of the dot-com bubble. Although the hedge fund industry as a whole was not affected in a major way, there were some famous casualties.[29] In contrast, investors' appetite for hedge funds improved and the industry saw a net asset inflow of 19.82% over this period.[30] This turn-around in the demand growth for hedge funds coincided with a major shift in the structure of the hedge fund industry. Figure 9.2 shows that for the first time since 1996, the cumulative performance of hedge funds (as measured by the DJCSI and the HFRI) exceeded that of the SNP index.[31] Not only had hedge funds outperformed the SNP index, their return standard deviations were just over half of the SNP index's return standard deviation. This caught the attention of another major group of investors—institutional investors such as foundations, endowments, pension funds, and insurance companies. Figure 9.3 charts the allocation of assets from some of these institutions to hedge funds. The subsequent two-year periods of 2002–2003 and 2004–2005 both experienced over thirty percent increases in net asset growth for the hedge fund industry, according to the same Lipper-Tass Asset Flows Report. By the end of 2007, the total Assets-Under-Management ("AUM") of the hedge fund industry had grown to $1,390 billion from $197 billion at the end of 1999 according to the Lipper-Tass Asset Flow Report. Taken together these observations are consistent with a shift in the investor clientele from an industry dominated by private wealthy investors to institutional investors.

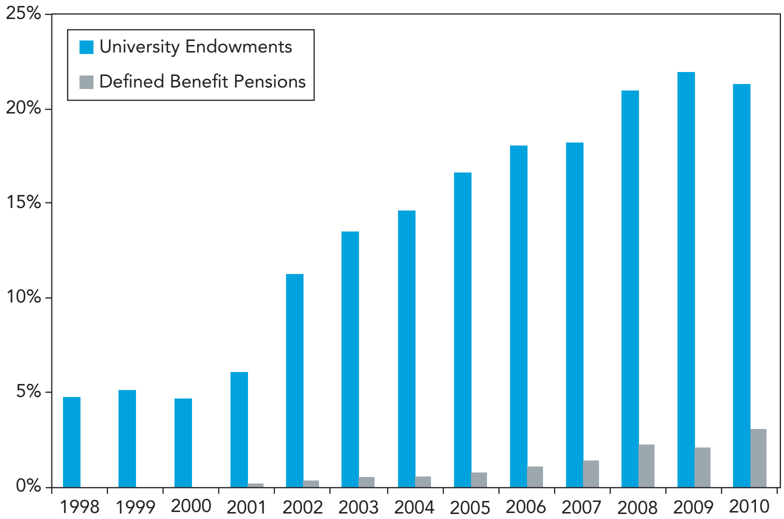


Figure 9.3 **Allocation to hedge funds (% of assets).**

Source: Pensions and Investments (1998–2010), NACUBO (1998–2008), NACUBO-Commonfond (2009–2010).

Hedge Fund Performance—The Post Dot-com Bubble Era

Institutional investors,[32] unlike private wealthy investors, tend to have moderate return expectations, lower tolerance of risk and above all even lower tolerance of high fees. Were institutional investors rewarded for allocating capital away from low-cost equity investments to hedge fund vehicles that demand much higher fees? Evidence from the eight year periods from 2002 to 2010 appears to favor such a move.[33]

Figure 9.4 plots the cumulative performance, starting in 2002, of the DJCSI, HFRI, and SNP index. Also added is the cumulative performance of the HFR Funds-of-Hedge Fund index ("HFRFOFI" for short). During the early days of institutional investments into hedge funds, it was common to engage an intermediary such as a fund-of-hedge fund manager who is charged with the responsibility of constructing a diversified portfolio of hedge funds that meets the investment objective of the

[29] Notably the famous Tiger Fund managed by well-respected industry veteran Julian Robertson ceased operations just one month before the burst of the bubble, to be followed one month later by the departure of Stanley Druckenmiller, then CIO of Soros Asset Management. For further analysis of hedge fund involvements in the dot-com bubble see Brunmermeir and Nagel (2004).

[30] According to the Lipper-Tass Asset Flows Report.

[31] Over the period 1996–2001, the cumulative performance of DJCSI, HFRI and SNP index were respectively 207.20%, 209.30%, and 204.41%. The corresponding annualized standard deviations were 9.81%, 8.80%, and 16.91%.

[32] See http://www.pionline.com/specialreports/plan-sponsors/20110207 and http://www.nacubo.org/Research/NACUBO-Commonfund_Study_of_Endowments/Public_NCSE_Tables.html for further reference.

[33] Our choice of 2002 as a starting point draws from two observations. Recall from Figure 9.2 that the SNP index began to underperform hedge fund indices towards the latter part of 2001. The capital allocation chart in Figure 9.3 also shows a rising pattern of increasing hedge fund allocations starting in 2002.

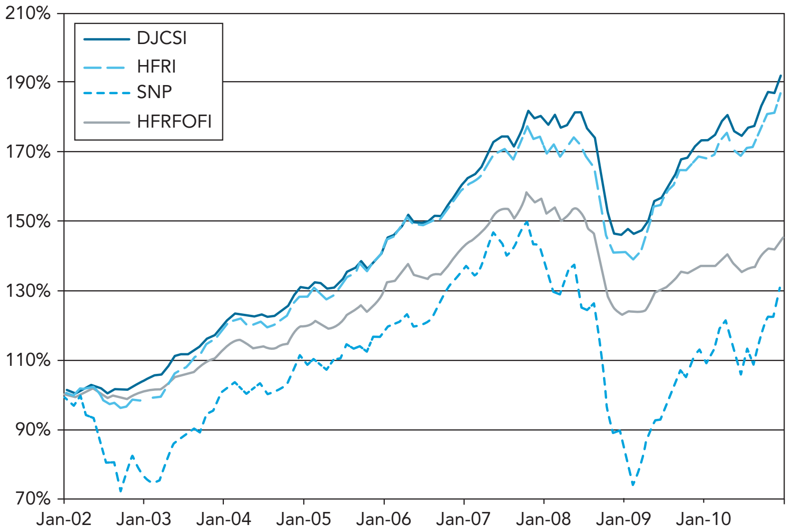


Figure 9.4 **Cumulative return of DJCSI, HFRI, HFRFOFI, and the SNP index: 2002–2010.**

investing institution.[34] Despite the additional layer of fees charged by funds-of-hedge funds, the cumulative net (of all fees and expenses) performance of hedge funds did much better than the SNP index. Cumulative performances over this period (2002–2010) were 72.64%, 69.82%, and 38.18% for the DJCSI, HFRI and HFRFOFI respectively. In contrast, the SNP index returned only 13.50%. The annualized standard deviations of returns are 5.84%, 6.47%, 5.51%, and 16%, respectively for the DJCSI, HFRI, HFRFOFI and SNP index.[35] Up to this point, empirical evidence suggests that switching from equity investments to hedge funds from 2002 delivered encouraging results. In the next few sections we explore the questions of performance persistency and omitted risk factors in order to better assess future prospects of this asset allocation decision.

Absolute Return and Alpha—A Rose by Any Other Name?

Up to this point we have traced the performance and capital formation of an opaque hedge fund industry from the investors' vantage point. We argued that the arrival of commercially available databases of hedge funds dramatically altered the flow of information between hedge funds and their investors. Together with adverse market events in 1994 and 1998, these changes profoundly impacted investors' perception of hedge fund investments. The long held belief among early investors of the hedge fund industry that skillful hedge fund managers can navigate their way safely through almost any adverse market-wide crisis was now seriously challenged. While a few managers did well during these market crises, the majority did not, and the statistics were now widely available with the arrival of commercial hedge fund databases. It is a rude awakening to find that just spreading one's capital among hedge fund managers engaging in different sounding strategies can expose investors to the same limited set of risk factors. Clearly, better understanding of the inherent risks of different hedge fund strategies is critical in constructing a diversified portfolio—an important topic we take up in the next section. In addition, these events may have also shaped the way hedge fund management companies develop their products; shifting from targeting out-sized returns from highly leveraged bets to emphasizing the value of survival and in turn risk management.

The turn of the century marked another major shift in the clientele makeup of the hedge fund industry—the emergence of institutional investors as the dominant investor group after the dot-com bubble. Not only are institutional investors different from private wealthy individuals, who dominated the early hedge fund industry, in their risk/return preferences, they are also much more demanding in operational integrity as well as the governance process in hedge fund firms. In other words, not only the quantity of hedge funds performance matters but the quality of performance also matters. Factors such as risk management, investment process, operational governance etc., all of which contribute to the key consideration of performance persistency, now become important. Some of these considerations are discussed later. An important consideration of institutional investors entering the hedge fund industry is how to benchmark their investment. A good investment benchmark should adequately capture past performance characteristics so as to guide investors on future return expectation and the attendant risk of the investment. We conclude this section on historical hedge fund performance with a discussion on benchmarking hedge fund strategies.

[34] In effect the investing institution not only out-sourced the portfolio construction, performance reporting and risk management of a hedge fund portfolio but also the day-to-day operational burden of managing a diversified portfolio of hedge funds which frequently runs upwards of 20 different hedge fund managers engaging a variety of investment strategies operating from geographical diverse locations.

[35] It should be noted that the outperformance of the SNP index by these hedge fund indices is driven primarily by two negative years—2002 and 2008.

Not all institutional investors came to the hedge fund industry taking refuge from the equity market. Some, like early investors in hedge funds, are looking for absolute performance (return) while others may view hedge funds as alternative sources of return with respect to a broader reference portfolio mix than just equities. What is common across most institutional investors is the need to determine whether, and how much, to allocate to the various hedge fund strategies, and whether these strategies deliver uncorrelated returns to a given reference portfolio mix (of conventional assets).

While strategy benchmarks based on peer-group averages can help investors compare returns of a given fund to its peers—peer group alpha—their short histories make it impossible for investors to evaluate performance behavior over different economic cycles. Put differently, how do peer group alphas behave during different market environments? Take, for example, the HFR Fixed Income Arbitrage index whose return history started in 1990. Set aside for the moment the data concerns prior to 1996. Suppose an investor in the mid 2000s, prior to the financial crisis of 2008, wanted to know how that strategy would behave if credit spread expands dramatically like in the 1930s. There is no direct way to answer that question, since there were no known hedge funds operating back in the 1930s. Yet this is a key consideration for investors whose reference portfolio is significantly exposed to high-yield bonds. Here, peer group alpha offers little help. Knowing that a particular fixed income arbitrage manager will outperform his or her peers may be of little solace if fixed income arbitrage as a strategy suffers large losses in a credit crunch.

During the second half of the 1990s, researchers tried to link hedge fund returns to market risk factors that often have a much longer performance history than hedge funds. The intuition behind such an approach can be illustrated by the following example. Suppose we are able to establish a stable relationship between the returns of the Fixed Income Arbitrage strategy to credit spread based on observed data (say from 1990 until 2007) like a credit spread beta. Investors with significant exposures to credit risk in their reference portfolio can now apply their outlook on the credit market to form consistent expectations for the performance of fixed income arbitrage hedge funds. Therefore, despite the lack of direct performance history that tells us how fixed income arbitrage hedge funds would have done during a credit crisis like the 1930s, knowing the strategy's credit spread beta could have forewarned investors how a 2008-style credit crisis would impact this strategy's performance. Applying the same logic, an investor would also be able to estimate how much of a fixed income arbitrage hedge fund's performance is persistently unrelated to the main risk driver of this strategy—factor-based alpha.

While there is no objective definition of absolute return from risky investments, a generally accepted description of absolute return is one that is non-negative irrespective of market conditions—in both bull and bear markets. An immediate question that arises is what one means by market—is it stocks, bonds, commodities or some other asset category? In other words, the concept of absolute return in and of itself implies a set of risk factors to which performance is independent. Therefore, given a set of risk factors and a performance evaluation period, there are identifiable similarities between the concept of absolute return and alpha. Obviously, if factor-based alpha can be separated from a strategy's factor beta through hedging techniques, it would have much of the attributes of an absolute return investment. However, the reality is hedge fund products are typically packaged with both components—alpha and the attendant risk factor exposures that drive a strategy's performance. This raises the natural question: how much systemic risk (risk factors) is (are) there in a diversified portfolio of hedge funds and do these risk factors interact with conventional asset classes?

We defer a fuller description of hedge fund exposure to market risk factors and the question of alpha-beta separation in hedge fund investing to later sections. As an illustration, we make use of a set of risk factors commonly used in hedge fund empirical research. Our set consists of eight risk factors. Seven of these were first proposed in Fung and Hsieh (2004b). Since then, we have added an emerging market equity factor, after the importance of these new markets started to show up in hedge fund returns in 2005.[36] In the eight-factor model, there are three equity factors: the excess return of the S&P 500 index over the risk-free return as proxied by the three-month T-bill (SP-Rf), the return of small cap stocks as proxied by the Russell 2000 index in excess of the S&P 500 index (RL-SP), and the excess return of the IFC Emerging Market Index over the three-month T-bill (IFC-Rf). These are typically risk factors found in equity strategies, such as Long/Short Equities and Emerging Markets. There are two bond factors: the excess return of the ten-year Treasury Note over the three-month bill (TY-Rf) and the return of Moody's BAA bonds over the return of ten-year Treasury Note (BAA-TY). These bond factors are found in various bond strategies, including Fixed Income Arbitrage, Distressed Securities, as well as Global/Macro funds that bet on interest rate policies in different countries. Lastly, there are three option portfolios: the excess returns of bond straddles, currency straddles, and commodity straddles, over the three-month T-bill (respectively, PTFSBD-Rf, PTFSFX-Rf, and PTFSCOM-Rf). These option returns mimic the

[36] See Edelman et al. (2012) for an empirical support for the presence of this eighth factor.

"long volatility" behavior of trend-following Managed Futures funds and Global/Macro funds. They also pick up "short volatility" characteristics of arbitrage-type strategies.

Table 9.2 summarizes the results from regressions of the returns of three hedge fund benchmarks on these eight risk factors.

Over the entire sample period, Table 9.2A shows that the intercept term (alpha relative to the eight-factor model) is highly statistically significant for both indices of the hedge fund industry—HFRI and DJCSI, respectively 26 and 29 bps (basis points) per month or 3.12% and 3.38% per annum. The intercept term (or factor-based alpha) for the index of funds-of-hedge funds (HFRFOFI) is statistically not different from zero. This is consistent with the additional layer of fees levied by funds-of-hedge fund managers. Put differently, over the entire sample period, FOHFs may have delivered some factor-based alpha but only sufficient to pay for the additional expenses of their services. Table 9.2a also shows significant exposures to the equity risk factors (SP-Rf and RL-SP). In addition, exposures to illiquid risk factors such as credit spread (BAA-TY) and emerging market equities (IFC-Rf) are statistically significant. On average, there is no statistically significant exposure to the dynamic trend-following factors over the entire sample period.

But is this a consistent description of hedge fund performance in sub-periods leading up to and after major market events that we have identified? To examine potential changes in performance in the hedge fund industry over time, Table 9.2B–D contain the factor regression over three sub-periods: 1996–1998, 1999–2001, and 2002–2010.

In terms of alpha relative to the eight-factor model, this was positive in the first subperiod (1996–1998) but only statistically different from zero for the HFRI (at 0.61% per month). Alpha was uniformly positive and statistically significant for all three benchmarks in the second sub-period (1999–2001), being 0.42%, 0.67%, and 0.44%, respectively, for DJCSI, HFRI, and HFRFOFI. In contrast, alpha is close to zero and not statistically significant in the last sub-period (2002–2010) for all three benchmarks. Prima facie one may be tempted to conclude that there may be some factor-based alpha during the pre-dotcom bubble period, but certainly, on average at the index level, these alphas have all but dissipated. However, more recent academic research has generally concluded that there are significant biases in observed returns from databases. The earlier the data, the less reliable they are and the more biases they contain—survivorship, backfilled, and self-selection. Therefore, pre-the LTCM crisis, performance data were "sampled" mostly from hedge funds that were eager to disclose their "good" performance (self-selection) which may have been extrapolated from a successful "incubation period" with very few investors, limited capital in the fund, leaving out failed funds that never reported. Consequently one has to be cautious on reported alphas from early data.[37] The uniformly large and statistically significant alpha during the post-LTCM recovery period from all three indices of hedge fund performance must also attract similar skepticism. However, it must be noted that data quality post-1998 has been on the rise. Therefore, these reported alphas from the second sub-period may well be a consequence of hedge fund managers reaping the rewards of the favorable market conditions from the wholesale liquidation of risky assets from the LTCM debacle.[38] Be that as it may, factor-based alphas that spike and ebb may tell us something about the volatility of market dynamics, but they are hardly confirmations of persistent absolute-return-like performance characteristics. If one were to interpret hedge fund alphas as persistent, stable, absolute returns not driven by systemic events of the market, then the preponderance of empirical data does not favor their presence in observable history—at least at the industry average or index level.

What kind of performance characteristics do hedge funds offer that justify the growing presence of institutional investors? This subject will be taken up in more detail later. For now, we need to develop more tools to help us understand the inherent risks in hedge fund investing.

In order to gain some broad insight into the type of risks that drove historical hedge fund performance, we continue with analyzing the results in Table 9.2. In terms of exposures to the equity factors, all three benchmarks had positive and statistically significant exposure to the SNP index in the first two sub-periods, dropping substantially in the last sub-period—only the HFRI has statistically significant exposure. Exposure to the spread between small cap and large cap stocks (RL-SP) was positive in the first sub-period, with HFRI and HFRFOFI having statistically significant exposure. It rose in the second subperiod, being statistically significant for all three benchmarks. As in the case of the SNP index, the exposure to RL-SP also became statistically insignificant in the third sub-period. In terms of exposure to emerging market equities, the exposure rose steadily across the three sub-periods, and they are all statistically significant in the second and third sub-periods.

With respect to fixed income factors, only the DJCSI showed statistically positive exposure to US treasuries (TY-Rf) in the

[37] Certainly the fact that funds from one index provider, HFRI, show a large statistically significant alpha (0.61% per month), while another index provider's data, DJCSI, shows no statistically significant factor alpha does not inspire confidence. Another issue of concern is the differences in the way these indices are constructed. See Table 9.2B.

[38] Similar performance spikes can also be found post 2008 as markets recover.

Table 9.2 Regression of Hedge Fund Indices on Eight Risk Factors: SP-Rf is the Excess Return of the S&P 500 Index. RL-SP is the Return of the Russell 2000 Index Minus the Return of the S&P500 Index. TY-Rf is the Excess Return of US Ten-Year Treasuries. BAA-TY is the Return of Moody's BAA Corporate Bonds Minus the Return of US Ten-Year Treasuries. PTFSBD-Rf is the Excess Return of a Portfolio of Bond Straddles. PTFSFX-Rf is the Excess Return of a Portfolio of FX Straddles. PTFSCOM-Rf is the Excess Return of a Portfolio of Commodity Straddles. IFC-Rf is the Excess Return of the International Finance Corporation's Emerging Market Index

2A	DJCSI	HFRI	HFRFOFI
Full sample: 1996–2010			
Const	0.0026	0.0029	0.0005
	0.0012	*0.0009*	*0.0012*
	2.1272	**3.0487**	**0.4451**
SP-Rf	0.1511	0.1884	0.0690
	0.0472	*0.0343*	*0.0337*
	3.2002	**5.4880**	**2.0463**
RL-SP	0.1083	0.1632	0.0870
	0.0661	*0.0389*	*0.0417*
	1.6388	4.1918	2.0889
TY-Rf	0.1321	0.0091	0.0309
	0.0719	*0.0459*	*0.0513*
	1.8375	0.1986	0.6019
BAA-TY	0.2034	0.0945	0.1597
	0.0563	*0.0342*	*0.0532*
	3.6154	**2.7667**	**3.0047**
PTFSBD-Rf	−0.0223	−0.0042	−0.0121
	0.0116	*0.0073*	*0.0098*
	−1.9205	−0.5786	−1.2359
PTFSFX-Rf	0.0068	0.0047	0.0051
	0.0070	*0.0041*	*0.0046*
	0.9730	1.1668	1.1104
PTFSCOM-Rf	0.0171	0.0043	0.0078
	0.0104	*0.0056*	*0.0073*
	1.6503	0.7751	1.0655
IFC-Rf	0.0797	0.1282	0.1207
	0.0244	*0.0171*	*0.0194*
	3.2729	**7.5163**	**6.2172**
Adj R^2	0.4666	0.7838	0.5713

2E	DJCSI	HFRI	HFRFOFI
p-value equality of coefficients			
Const	0.0030	0.0000	0.0002
SP-Rf	0.0447	0.0438	0.0461
RL-SP	0.0001	0.0001	0.0238
TY-Rf	0.0241	0.0398	0.0554
BAA-TY	0.0122	0.0001	0.0049
PTFSBD-Rf	0.8092	0.2472	0.9943
PTFSFX-Rf	0.6158	0.5558	0.6646
PTFSCOM-Rf	0.5192	0.7368	0.6023
IFC-Rf	0.5833	0.3116	0.4724
All coefficients	0.0000	0.0000	0.0000

Standard errors, using Newey and West (1987) with 6 lags, in italics.

	DJCSI	HFRI	HFRFOFI	DJCSI	HFRI	HFRFOFI	DJCSI	HFRI	HFRFOFI
	2B. Subsample: 1996–1998			*2C. Subsample: 1999–2001*			*2D. Subsample: 2002–2010*		
Const	0.0028	0.0061	0.0015	0.0042	0.0067	0.0044	0.0007	0.0006	−0.0017
	0.0043	*0.0016*	*0.0027*	*0.0021*	*0.0017*	*0.0021*	*0.0013*	*0.0008*	*0.0011*
	0.6512	3.8125	0.5556	2.0000	3.9412	2.0952	0.5385	0.7500	−1.5455
SP-Rf	0.2558	0.2695	0.1598	0.1992	0.3058	0.1187	0.0208	0.0708	−0.0273
	0.0866	*0.0329*	*0.035*	*0.0492*	*0.0303*	*0.0325*	*0.0366*	*0.0266*	*0.0314*
	2.9538	8.1915	4.5657	4.0488	10.0924	3.6523	0.5683	2.6617	−0.8694
RL-SP	0.0925	0.2903	0.155	0.2601	0.2229	0.1544	0.0117	0.0564	0.0033
	0.1106	*0.0348*	*0.0567*	*0.0379*	*0.0334*	*0.0371*	*0.0447*	*0.0438*	*0.0461*
	0.8363	8.3420	2.7337	6.8628	6.6737	4.1617	0.2617	1.2877	0.0716
TY-Rf	0.5951	0.071	0.197	0.1765	0.1438	0.138	−0.005	−0.0437	−0.0316
	0.2445	*0.058*	*0.1068*	*0.1087*	*0.068*	*0.0907*	*0.0571*	*0.0406*	*0.0451*
	2.4339	1.2241	1.8446	1.6237	2.1147	1.5215	−0.0876	−1.0764	−0.7007
BAA-TY	1.2684	0.7318	1.0851	−0.011	−0.0135	0.0142	0.1962	0.142	0.1662
	0.3698	*0.14*	*0.2886*	*0.1807*	*0.0816*	*0.1112*	*0.0544*	*0.032*	*0.05*
	3.4300	5.2271	3.7599	−0.0609	−0.1654	0.1277	3.6066	4.4375	3.3240
PTFSBD-Rf	−0.037	0.0042	−0.0093	−0.0181	−0.015	−0.0106	−0.0143	−0.003	−0.0115
	0.0353	*0.013*	*0.0259*	*0.017*	*0.0057*	*0.0095*	*0.0097*	*0.0055*	*0.0082*
	−1.0482	0.3231	−0.3591	−1.0647	−2.6316	−1.1158	−1.4742	−0.5455	−1.4024
PTFSFX-Rf	−0.0056	0.0055	0.0027	0.0166	0.0164	0.0158	0.0077	0.005	0.0055
	0.0206	*0.0054*	*0.01*	*0.0115*	*0.0101*	*0.0119*	*0.0043*	*0.0031*	*0.0035*
	−0.2718	1.0185	0.2700	1.4435	1.6238	1.3277	1.7907	1.6129	1.5714
PTFSCOM-Rf	0.0455	0.0025	0.0243	0.003	0.0113	0.0068	0.0018	0.0007	−0.0007
	0.0375	*0.0142*	*0.0253*	*0.0171*	*0.0123*	*0.0139*	*0.0073*	*0.0057*	*0.0072*
	1.2133	0.1761	0.9605	0.1754	0.9187	0.4892	0.2466	0.1228	−0.0972
IFC-Rf	0.0609	0.1189	0.0985	0.0992	0.1529	0.1582	0.1397	0.1726	0.163
	0.0876	*0.0322*	*0.0525*	*0.0493*	*0.0242*	*0.0357*	*0.019*	*0.015*	*0.0241*
	0.6952	3.6925	1.8762	2.0122	6.3182	4.4314	7.3526	11.5067	6.7635
Adj R^2	0.4799	0.9079	0.6879	0.5841	0.8616	0.6495	0.6452	0.8259	0.6361

first sub-period. All three benchmarks had statistically positive exposure to TY-Rf in the second sub-period, while becoming insignificant in the third sub-period. Exposures to the credit spread (BAA-TY) are positive and statistically significant for all three benchmarks in the first and third sub-periods, but not in the second sub-period. In terms of exposure to the three volatility factors (PTFSBD, PTFSFX, and PTFSCOM), the three benchmarks did not show statistically significant exposures. The main reason for this result is that some strategies have positive exposures (e.g. Managed Futures) and some have negative exposures (e.g. Distressed). In Table 9.2E, we test for the equality of the exposures in the three sub-periods. At the 5% significance level, the null hypothesis of equality is not rejected for the three volatility factors and emerging market equities, but is rejected for the other equity factors and the two fixed income factors.

Taken together, these results suggest that both the exit of wealthy individuals and the entry of institutional investors appear to be justified but for different reasons. Early investors did benefit from a higher average return. However, this is not readily separable from the attendant exposures to a number of other risk factors. These additional risk exposures cost investors dearly when the LTCM saga unfolded. In short, there may be alpha but there is also more risk and, unlike conventional asset classes which are liquid and readily executable, alpha from a hedge fund portfolio often comes with inseparable systemic exposures to risk factors, some of which may be difficult to implement. To what extent these risk factor exposures can be managed to allow investors to extract the alpha component of total performance is a question we defer to later. Suffice to say that for institutional investors entering the hedge fund market in search of alternative sources of return away from the equity market, the results over the later period of 2002–2010 are encouraging. On average, there is very low exposure to the equity risk factors from hedge funds. However, at the index level, returns were earned as a premium for bearing other factor risks such as credit and emerging markets. In short, there may not be alpha in the conventional sense of securities selection skills, but there appears to be return from bearing an alternative source of risk to conventional equities.[39]

Comparing the above results to the regression model's output for the entire 1996–2010 period, the evidence suggests that returns from the three sub-periods appear to be drawn from different return distributions. This is consistent with the results in Fung et. al. (2008) and Kosowski, Naik, and Teo (2007). The decline in the intercept term of the regression model, both numerically and in statistical significance, is also consistent with a maturing hedge fund industry which has been growing steadily towards the point of exceeding their capacity to generate abnormal returns adjusting for risk—an outcome that is anticipated by the theoretical model of active management by Berk and Green (2004). The approach here is to qualitatively identify discrete, demand-driven turning points (in the sense of investor preferences) in the hedge fund industry and examine how this may affect risk taking behavior of hedge funds. We posit that these demand-driven structural changes ultimately translate to observable changes in the performance characteristics.[40] It is comforting that the overall conclusions reached with the qualitative, demand-driven approach are consistent and reconcilable to those reached using statistical analysis of hedge fund returns.[41]

The results thus far add color to the overall hedge fund landscape, albeit with broad strokes, and highlights the important but changing role investors play in shaping the hedge fund industry. Going forward, how would the capital and product formation of the hedge fund industry change? We can break down this question into its components. First, in the absence of obviously identifiable alpha, will institutional investors continue to tolerate the high cost (fees) of hedge fund investing? Second, if hedge funds do not deliver alpha, what do they deliver to investors? Thus far, our analysis followed a top-down approach and analyzed the hedge fund industry at the index level. However, one of the major attractions of the hedge fund industry is its rich tapestry of strategies with diverse performance characteristics. As analytical tools and data become more readily available, one has to expect portfolio construction processes to morph in sophistication. It would be reasonable to ask the additional questions: third, does manager selection matter (or forget the average, let me pick the best of the breed)? Fourth, how do successful hedge fund managers differentiate themselves from their peers? For instance, are there successful large hedge funds delivering stellar performance like those we saw in Figure 9.1 during the pre-1996 era of the hedge fund industry? Fifth, should investors be concerned about *Black Swans* lurking in their portfolio? These questions will be analyzed after we summarize the development of portfolio management tools from the past decade of hedge fund research.

[39] Interpreting alpha in the spirit of Sharpe (1992).

[40] In a sense, this is similar to changes in investment behavior of mutual fund managers in respond to changes in their benchmark (or investment mandate). The difficulty with hedge funds is the lack of a recorded history of explicit, contract-like mandates from investors—hence the statistical detective work.

[41] Statistical models on regime change/sample break based on observed hedge fund returns essentially look for switch points based on supply-side (hedge fund managers) behavior.

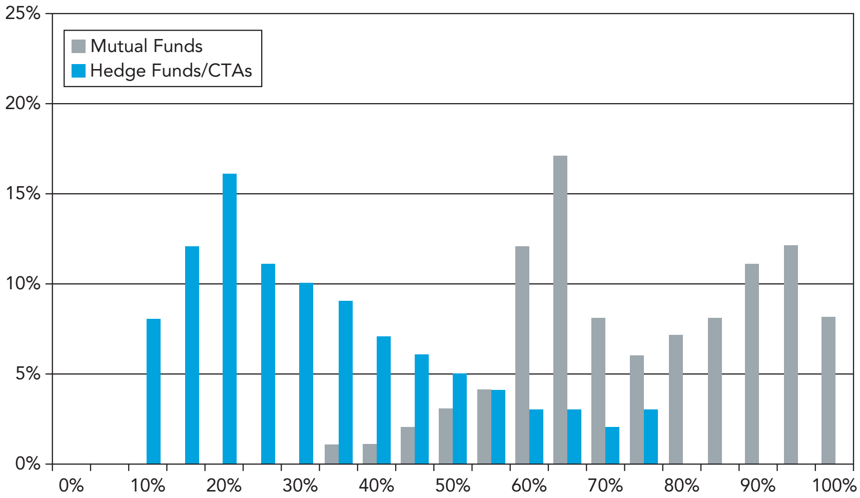


Figure 9.5 **Distribution of R^2s versus asset classes.**

Source: Fung and Hsieh, 1997, Figure 1.

9.3 THE RISK IN HEDGE FUND STRATEGIES

From Passive Index Strategies to Active Hedge Fund Styles

The return characteristics of a passive buy-and-hold strategy applied to a conventional asset class, by design, mirror the characteristics of that asset class's returns. Sharpe (1992) introduced the idea of a *style model* to capture active management strategies applied to a conventional asset classes. In Sharpe's model, an investment style is extended to include linear combinations of an expanded set of asset-class indices—extended to allow for sector specialization. In this formulation, investment styles differ from each other by the choice of asset-class indices and the exposure (β) to each index—market leverage. In essence, Sharpe's model depicts style in two dimensions—the choice of asset class indices, and the level of β. Since asset class indices typically assume a passive long-only, buy-and-hold strategy, the activity of Sharpe's model is limited to choosing the size of bet (β) applied to each of a selected set of passive strategies (choice of asset class indices).

Fung and Hsieh (1997a) extend Sharpe's model to hedge fund styles. First, the investment opportunity set is extended beyond conventional asset classes to reflect hedge fund managers' ability to engage in outright short sale of securities—thus allowing betas to be negative.[42] Second, exposure to a given asset class is extended to reflect the dynamic nature of hedge fund strategies—allowing betas to vary over time. Third, the range of *asset-class betas* has to be broadened to allow for the use of financial leverage as well as market leverage—*financial betas versus market betas.*[43] In the Fung and Hsieh (1997a) framework, hedge fund styles can be thought of as pair-wise combinations of an extended opportunity set and strategies that admit dynamic betas as well as leverage. Figure 9.5 contrasts the low correlations between hedge fund returns and asset-class indices to the high correlations between mutual fund returns and the same set of asset-class indices.

Figure 9.5 is reproduced from Figure 9.1 of Fung and Hsieh (1997a). It represents the distribution of R^2s of regressions of hedge fund returns and mutual fund returns on eight asset classes (US equities, non-US equities, emerging market equities, US government bonds, non-US government bonds, one-month Eurodollar deposit rate, gold, trade-weighted value of the Dollar). While 48% of hedge funds have R^2s below 0.25, 47% of mutual funds have R^2s above 0.75. This indicates that hedge fund returns have low correlation with standard asset returns, quite different from mutual fund returns. Consistent with Sharpe (1992), mutual fund returns tend to be highly correlated to different asset class indices depending on their style. In the Fung and Hsieh (1997a) framework, the low correlation between hedge fund returns and asset class indices can come from two sources—transacting in securities not correlated to conventional asset class indices or the dynamic nature of their trading strategies.

To identify the systematic factors that motivate hedge fund returns we need to tackle both performance dimensions—we need to identify the markets in which hedge funds transact (*locate where* hedge funds trade) and establish a set of transparent, rule-based strategies that mimic the hedge fund strategies (*model how* hedge funds trade). The development of a complete model of hedge fund styles has followed a somewhat chaotic path.

[42] Beyond the under weighting or over weighting of a security in a given index—or tilts.

[43] Total exposure to a given asset class is allowed to exceed one.

Peer-Group Style Factors

To help investors understand hedge funds, consultants and database vendors categorize hedge funds based on individual manager's description of the strategies each fund uses. Averaging the funds' returns in each group gives the style index.[44] This type of qualitatively determined peer-group averages remains the most common hedge fund index construction method to date. While peer-group averages are transparent in their construction and easy to understand, other ambiguities can arise. For instance there is no generally agreed standard for classifying hedge funds into homogeneous groups based on self-description.[45] Over time, hedge fund managers have to grow and evolve their business to cope with changing market conditions—in short, styles can change and do.[46] Indeed, some index providers, such as HFR, have changed their indices several times over the past decade. Above all, there is no obvious way to link qualitative styles indices to quantifiable market factors. This in turn limits the insight on the underlying hedge fund strategies static style indices can provide.[47]

Return-Based Style Factors

To address some of these issues, Fung and Hsieh (1997a) proposed a complementary quantitative approach. Instead of relying on hedge fund managers' self-disclosed description of their strategies, one can construct groups of hedge funds that exhibit similar return characteristics. In other words, *don't just rely on what hedge fund managers say they do, look also at what they actually do.*

The Fung and Hsieh (1997a) approach is predicated on the idea that funds using the same strategy should deliver more highly correlated returns than those using different strategies; this can be identified by the principal components of their historical returns. This choice of method is motivated by three reasons. First, statistical clustering of returns should approximate the common risk-return characteristics of the strategies they use and the markets in which they transact. Second, principal component analysis helps to reduce the myriad of peer-group-based style factors to a more manageable set. Third, qualitative, self-disclosed strategy information from hedge fund managers can be used as an exogenous source of data to identify these statistically constructed components which can be interpreted as return-based style factors.

Following this process, Fung and Hsieh (1997a) identified five return-based style factors. Since then, Brown and Goetzmann (2003) applied a variation of this approach on an updated dataset and found eight style factors. They interpreted these factors to be Global Macro (similar to Fung and Hsieh (1997a)); Pure Leveraged Currency (similar to the trend-following factor of Fung and Hsieh (1997a)); two equity factors—a US and a non-US factor (similar to the Value factor of Fung and Hsieh (1997a)); an Event-Driven factor (similar to the Distressed Factor of Fung and Hsieh (1997a)); and two sector specific factors—Emerging Markets and Pure Property (both excluded from the Fung and Hsieh (1997a) study). Like Brown and Goetzmann (2003), other studies of return-based style factors have generally identified similar factors which add credence to the proposition that there are only a limited number of systematic hedge fund strategies, or style factors, that persist over time. Unusual return characteristics from a number of these style factors help to explain the low correlation of hedge fund returns with conventional asset classes that we see in Figure 9.5. For example, Fung and Hsieh (1997a) reported a convex return profile of the Systematic/trend following factor with respect to US equity market performance as depicted in Figure 9.6.

Figure 9.6 plots the returns of the systematic/trend following factor's return conditional on different states of the US equity market. It shows this factor's return behaves in a nonlinear fashion depending on the condition of the US equity market—inversely correlated during down markets, seemingly uncorrelated during normal equity market conditions, rising to positive correlation during up market conditions. In short, it exhibits the return characteristics of a straddle on the equity market. A similar nonlinear, option-like, return characteristic from Global Macro hedge funds was also reported in Fung and Hsieh (1997a, p. 290).

It is this type of nonlinear return characteristic that makes hedge funds attractive for diversifying investments to conventional asset class portfolios. However insight on the drivers of these nonlinear, state-dependent returns behavior beyond just an

[44] The averaging process can be equally weighted or value (AUM) weighted. Different weighting schemes implicitly assume different portfolio strategies, some of which may not be compatible with the terms and conditions of the fund.

[45] Over the years, there appears to be a convergence of opinion among different hedge fund index providers on this subject, but not completely. To date, different hedge fund index providers may have similar sounding style indices with highly correlated returns, but complete convergence is still a long way away.

[46] For instance, at one point, HFR reports over 15 categories of hedge fund sub-indices reflecting different styles, whereas DJCS has generally maintained mostly the same set of style indices (under ten) throughout its history.

[47] This limits the risk metrics of qualitative hedge fund styles to statistical moments of historical returns.

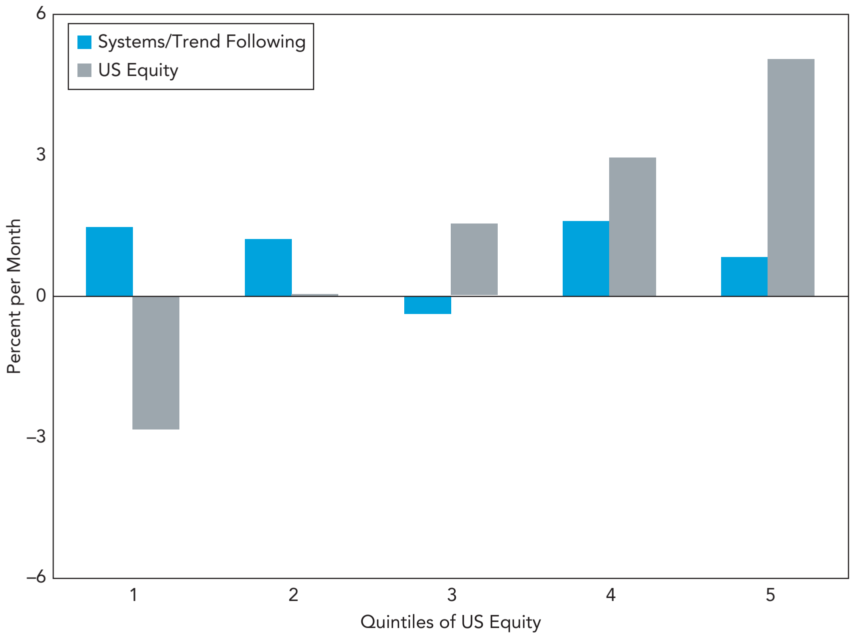


Figure 9.6 **Systems/trend following style versus US equity.**

Source: Fung and Hsieh (1997a) Figure 3.

empirical regularity is needed. We begin by addressing the question: *how do return-based factors relate to market prices?*

Top-Down versus Bottom-Up Models of Hedge Fund Strategy Risk

For the industry as a whole, return-based style factors help us identify the main markets in which hedge fund managers transact. However, to better understand the dynamics of how bets are determined for different strategies, we need a more micro (or bottom-up) approach to modeling the return generating process of different hedge fund strategies so as to explicitly identify the market factors that drive performance.

Research dedicated to identifying the risk factors inherent in specific hedge fund styles began with Fung and Hsieh (2001) and Mitchell and Pulvino (2001). Fung and Hsieh (2001) modeled the return of managed futures (trend followers) using option straddles, and Mitchell and Pulvino (2001) mimic the returns of merger arbitrage funds to a rule-based passive merger arbitrage strategy. Since then a number of other studies have modeled risk factors in other hedge fund styles—Fung and Hsieh (2002, 2006), Duarte, Longstaff, and Yu (2007), Brave et. al. (2008), Patton (2009), Agarwal et. al. (2010), Fung and Hsieh (2011).

Directional Hedge Fund Styles: Trend Followers and Global Macro

The majority of *managed futures funds* employs a *trend-following strategy*. Here's a description of the strategy that funds included in the DJCS Managed Futures index follow:[48]

> The Dow Jones Credit Suisse Core Managed Futures Hedge Fund Index^{SM} is a subset of the Dow Jones Credit Suisse Core Hedge Fund Index^{SM} that seeks to measure the aggregate performance of managed futures funds. Managed futures funds (often referred to as CTAs or Commodity Trading Advisors) typically focus on investing in listed bond, equity, commodity futures and currency markets, globally. Managed futures fund managers tend to employ systematic trading programs that largely rely upon historical price data and market trends. A significant amount of leverage is often employed since the strategy involves the use of futures contracts. CTAs tend not to have a particular bias towards being net long or net short in any particular market.

Fung and Hsieh (2001) show that majority of managed futures funds pursue trend following strategies. Merton (1981) showed that a market timer, who switches between stocks and Treasury bills, generates a return profile similar to that of a call option on the market. Fung and Hsieh (2001) generalized this observation to encompass both long and short positions. The resulting return profile is similar to that of a straddle. Over any observation period, a trend follower with perfect foresight would have been able to initiate (and to exit) trading positions at the best possible price. The payout function of a trend follower with perfect foresight therefore resembles that of a lookback straddle.[49] Since

[48] For more information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?cy=USD&indexname=CORE_MGFUT.

[49] A lookback straddle consists of a lookback call option and a lookback put option. A lookback call option allows the owner to buy the underlying asset at the lowest price during the life of the call option. A lookback put option allows the owner to sell the underlying asset at the highest price during the life of the put option. The lookback straddle therefore allows the owner to buy at the low and sell at the high. The lookback option was originally analyzed in Goldman, Sosin, and Gatto (1979).

the payout of a lookback option is the same as that of a trend follower with perfect foresight, the execution cost of such a straddle (or the price of the lookback option) for a given trend follower can be interpreted as reflecting the cost of initiating (and exiting) trading positions at sub-optimal points. *How well do lookback straddles mimic the performance of trend following hedge funds?*

Using exchange-traded standard straddles in twenty-six markets, Fung and Hsieh (2001) replicated returns of lookback straddles and grouped them into five portfolios according to the underlying assets—stock indices, bond futures, interest rate futures, currency futures, and commodity futures. Empirical evidence shows that three option portfolios—bonds, currencies, and commodities—have strong correlations to trend followers' returns at a level well beyond previous results.[50] It is intuitively appealing that market volatility has been a key determinant of trend-following returns. Since the Fung and Hsieh (2001) study, the relationship they reported has continued to hold. Figure 9.7 compares the performance of the lookback portfolios to trend followers. The solid line represents the monthly returns of trend followers (based on the DJCS Managed Futures index). The out-of-sample return forecasts of trend followers as represented by the dotted line in Figure 9.7, are generated using the regression coefficients from the regression equation in Fung and Hsieh (2001), which was applied to data ending in 1997, and the realized values of the lookback portfolios starting in 1998. The forecast returns continued to track the actual returns of trend followers after the Fung and Hsieh (2001) study.

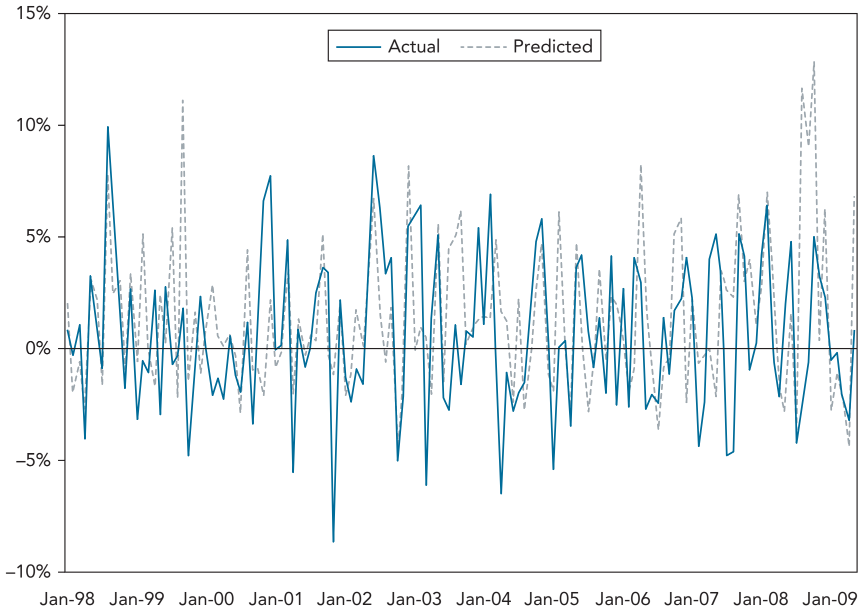


Figure 9.7 **Actual and predicted returns of DJCS managed futures index.**

Source: Fung and Hsieh (2006) Figure 3 updated.

Like most other qualitative description of hedge fund styles, there is no universally accepted definition of *Global Macro hedge funds* as a group. Below is how Global Macro managers are described in the Dow Jones Credit Suisse website:[51]

> Global macro funds typically focus on identifying extreme price valuations and leverage is often applied on the anticipated price movements in equity, currency, interest rate and commodity markets. Managers typically employ a top-down global approach to concentrate on forecasting how political trends and global macroeconomic events affect the valuation of financial instruments. Profits can be made by correctly anticipating price movements in global markets and having the flexibility to use a broad investment mandate, with the ability to hold positions in practically any market with any instrument. These approaches may be systematic trend following models, or discretionary.

Fung and Hsieh (2006) provide a detailed description of Global Macro hedge fund managers' performance. Global Macro fund managers are known to be highly dynamic traders, often taking highly leveraged bets on directional movements in exchange rates, interest rates, commodities, and stock indices in an

[50] Table 9.2 of Fung and Hsieh (2001) conventional commodity indices such as the Goldman Sachs Commodity Index and the Commodity Research Bureau Index offer little no explanatory power to trend following hedge funds' returns. The Mount Lucas/BARRA Trend-Following Index has better explanatory power at 7.5% adjusted *R*-square whereas the five portfolios of Lookback straddles has an adjusted *R*-square of 47.9%.

[51] For further reference, go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_GLMAC&cy=USD.

opportunistic manner. We can think of a Global Macro manager as a highly active asset allocator betting on a range of risk factors over time. Over a reasonably long time frame, opportunities in different risk factors will come and go. Ex post the performance of a Global Macro manager may well resemble that of a diversified hedge fund portfolio with highly variable exposures to a broad set of global risk factors. As a first approximation, Fung and Hsieh (2006) applied the seven factor model of Fung and Hsieh (2004b), which was originally designed to capture the inherent risks in diversified portfolios of hedge funds, and reported reasonable results. In this chapter, we update the results to the full eight-factor model.

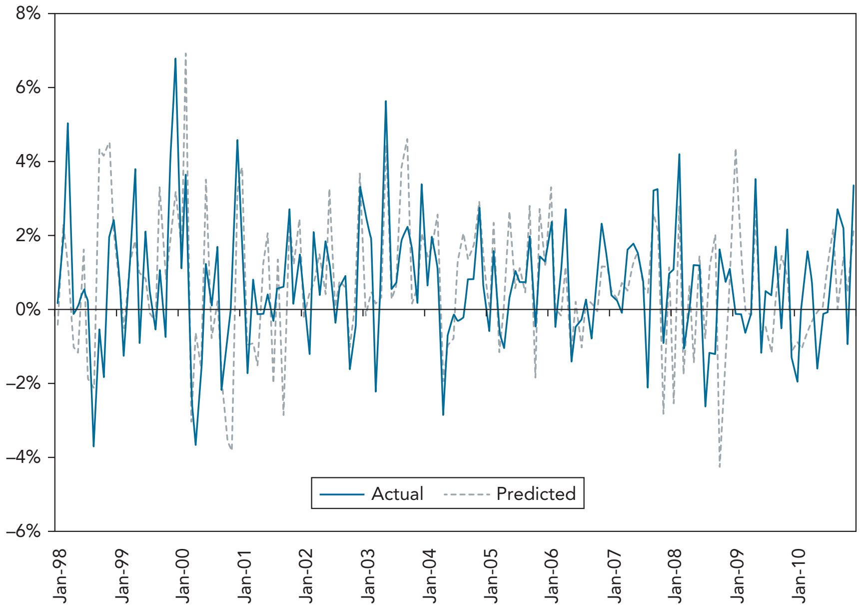


Figure 9.8 **Actual and predicted returns of HFRI global macro index.**

Source: Fung and Hsieh (2006) Figure 12 updated.

Figure 9.8 depicts the actual and one-month-ahead conditional forecast of the HFRI Global Macro Index.[52] For each month we use the prior twenty-four months' data to regress the macro index on the eight-factor model. Applying the regression coefficients and the realized values of the risk factors in the subsequent month, we generate the conditional one-month-ahead forecast giving the predicted value series in Figure 9.7. These results are consistent with the view that a Global Macro fund has similar return characteristics to a highly active portfolio of different hedge fund strategies. Therefore, tools that work well in capturing the risk characteristics of diversified hedge fund portfolios can be applied to describe the risk in Global Macro hedge funds.

An interesting similarity between Global Macro and Managed Futures is their "trend following" behavior. In the rolling regressions of the HFRI Global Macro index on the eight-factor model, the coefficient of the currency straddle portfolio is consistently positive—Global Macro funds have higher average returns during extreme moves in the currency markets. This is consistent with their behavior in Figure 9.5 of Fung and Hsieh (1997a, 1997b) which shows a nonlinear U-shaped pattern of Global Macro fund returns relative to changes in the Trade-Weighted Dollar Index.

Taking these results together, two points stand out in these two groups of directional trading styles. First is the similarity between Global Macro and Managed Futures as trend followers.[53] Second, these managers behave like asset allocators taking bets in different markets utilizing a range of strategies opportunistically. Together, these two points explain why they generate low return correlation to equities.

Event-Driven Hedge Fund Styles: Risk Arbitrage and Distressed

Sometimes also referred to as *Merger Arbitrage*, below is a description of this trading style from Dow Jones Credit Suisse's website:[54]

> Risk arbitrage event driven hedge funds typically attempt to capture the spreads in merger or acquisition transactions involving public companies after the terms of the transaction have been announced. The spread

[52] We prefer the Global Macro index from HFR over the one from DJCS, because the latter contained a very large fund that, on an AUM-weighted basis, dominated the return of that index.

[53] This view is consistent with our narrative on how the events of 1994 impacted the hedge fund industry. Quite simply, similarities between different hedge fund styles can lead to a convergence of risk exposure.

[54] For further references, go to: http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_MRARB&cy=USD.

is the difference between the transaction bid and the trading price. Typically, the target stock trades at a discount to the bid in order to account for the risk of the transaction failing to close. In a cash deal, the manager will typically purchase the stock of the target and tender it for the offer price at closing. In a fixed exchange ratio stock merger, one would go long the target stock and short the acquirer's stock according to the merger ratio, in order to isolate the spread and hedge out market risk. The principal risk is usually deal risk, should the deal fail to close.

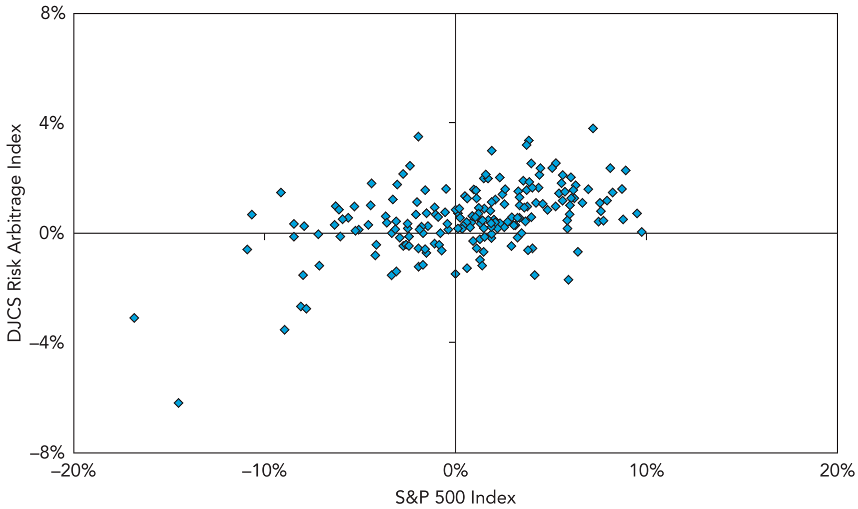


Figure 9.9 **Risk factor for DJCS risk arbitrage index: 1994–2010.**

Source: Fung and Hsieh (2006) Figure 4 updated.

Mitchell and Pulvino (2001) created an index of merger arbitrage returns, using announced mergers from 1964 until 2000. They showed that the merger arbitrage returns are similar to those of merger arbitrage hedge funds. In fact, they observe that both the merger arbitrage indices and merger arbitrage funds exhibit characteristics similar to a dynamically adjusted short position on the stock market. These results are illustrated in Figure 9.9, which plots the DJCSI Risk Arbitrage Index against the return of the S&P 500 index.

Note that the largest negative monthly returns from the DJCS Risk Arbitrage index all occur in months when the SNP index experienced a sizeable negative return. Essentially, merger arbitrageurs are betting on the consummation of a merger—in general, they are long "deal" risk. Their return can be viewed as the insurance premium from selling a policy against the failure to complete a merger. Typically, mergers fail for idiosyncratic reasons and can be diversified away in a portfolio of such transactions. However, when the stock market undergoes a severe decline, mergers tend to be called off for a variety of reasons—ranging from funding and pricing issues to concerns over the long-term prospects of the economy. This scenario is one in which there is a convergence of *deal-risk* that cannot be easily diversified.

Risk arbitrage is a sub-strategy category within the strategies utilized by hedge fund in the DJCS event-driven hedge fund style index. The other sub-strategy is generally referred to as Distressed Hedge Funds. Here is a qualitative description of Distressed hedge funds on the Dow Jones Credit Suisse's web site:[55]

> The Dow Jones Credit Suisse Event Driven Distressed Hedge Fund Index^{SM} is a subset of the Dow Jones Credit Suisse Hedge Fund Index^{SM} that measures the aggregate performance of event driven funds that focus on distressed situations. These funds typically invest across the capital structure of companies subject to financial or operational distress or bankruptcy proceedings. Such securities often trade at discounts to intrinsic value due to difficulties in assessing their proper value, lack of research coverage, or an inability of traditional investors to continue holding them. This strategy is generally long-biased in nature, but managers may take outright long, hedged or outright short positions. Distressed managers typically attempt to profit on the issuer's ability to improve its operation or the success of the bankruptcy process that ultimately leads to an exit strategy.

Since Distressed hedge funds can invest in a wide range of securities, many of which are not traded in the public market (e.g. bank loans, delisted equities, defaulted bonds), the main feature of this investment strategy is long exposure to credit risk

[55] For more information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_DISTR&cy=USD.

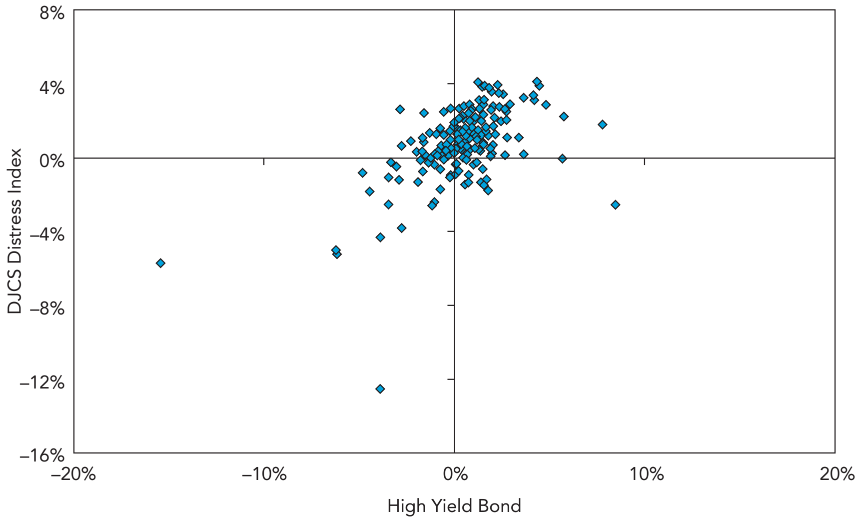


Figure 9.10 Risk factor in DJCS Distressed index: 1994–2010.

Source: Fung and Hsieh (2006) Figure 9 updated.

of corporations with very low credit ratings. While it is difficult to obtain returns of nonmarketable securities, publicly traded high-yield bonds are a good proxy for exposure to low grade credit. That turns out to be the case as shown in Figure 9.10.

The correlation between the DJCS Distressed index and high-yield bonds (as proxy by the Vanguard High Yield Corporate bond fund) is 0.55. There is, however, some evidence of nonlinearity between the two return series, particularly at the extreme tails. Since Distressed hedge funds own securities that are much less liquid than high-yield bonds, they may earn an extra liquidity premium over high-yield bonds, and they may also incur a higher funding cost for carrying very illiquid securities. While we have not found a good proxy for the funding cost of illiquid securities, we hypothesize that this cost would respond to extreme credit market conditions, particularly when short-term interest rates move dramatically. Indeed, a regression of the DJCS Distressed index on the returns of high-yield bonds and lookback straddles on the three-month Eurodollar deposit rate results in a positive coefficient to the former and a negative coefficient to the latter. This indicates that average returns of Distressed hedge funds are lower during extreme up and down moves in the three month Eurodollar deposit rate.

To summarize, the two major strategies in the Event-Driven hedge fund style category both exhibit nonlinear returns characteristics—mostly as tail risk that shows up under extreme market conditions. In the case of Risk Arbitrage, the tail risk is a large drop in equities. In the case of Distressed hedge funds, the tail risk is in a large move of short-term interest rates. However, unlike trend followers, who tend to benefit from extreme moves, Event-Driven funds are hurt by extreme moves.

Relative Value and Arbitrage-like Hedge Fund Styles: Fixed Income Arbitrage, Convertible Arbitrage, and Long/Short Equity

There are three main styles in this category separated by the market focus of the hedge fund managers. We begin with those managers with a fixed income market focus. DJCS reports an index of *Fixed Income Arbitrage Hedge Funds*. A description of the strategies utilized by funds included in this index is as follows:[56]

> The Dow Jones Credit Suisse Fixed Income Arbitrage Hedge Fund Index^{SM} is a subset of the Dow Jones Credit Suisse Hedge Fund Index^{SM} that measures the aggregate performance of fixed income arbitrage funds. Fixed income arbitrage funds typically attempt to generate profits by exploiting inefficiencies and price anomalies between related fixed income securities. Funds often seek to limit volatility by hedging out exposure to the market and interest rate risk. Strategies may include leveraging long and short positions in similar fixed income securities that are related either mathematically or economically. The sector includes credit yield curve relative value trading involving interest rate swaps, government securities and futures; volatility trading involving options; and mortgage-backed securities arbitrage (the mortgage-backed market is primarily US-based and over-the-counter).

In an academic study, Fung and Hsieh (2002) analyzed a broad sample of hedge funds from several commercial databases that are qualitatively classified as having a fixed income market

[56] For more information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=CORE_FIARB&cy=USD.

focus. For those hedge funds that are not classified with an arbitrage orientation, convertible bond funds were strongly correlated to the CSFB (Credit Suisse/First Boston) Convertible Bond Index. High-yield funds were strongly correlated to the CSFB High-Yield Bond index. In addition, all styles, including fixed income arbitrage, have correlations to changes in the default spread. Interpreting default spread as a measure of credit market condition, the central factor for most hedge funds with a fixed income market focus tend to be exposed to the credit market conditions. This is intuitively appealing as most hedge funds with a fixed income orientation depend on leverage to enhance performance.

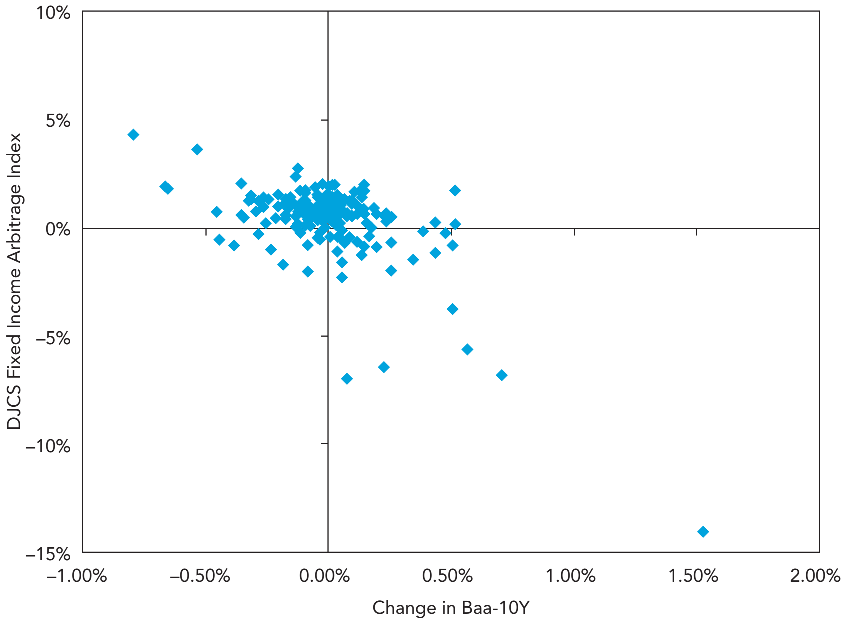


Figure 9.11 **Risk factor for DJCS fixed income arbitrage index: 1994–2010.**

Source: Fung and Hsieh (2006) Figure 5 updated.

Figure 9.11 provides support for this view. Here, we graph the then DJCS Fixed Income Arbitrage Index against the change in credit spread, as proxied by Moody's Baa yield over the ten-year Treasury constant maturity yield. A very similar picture is obtained if we use the HFRI Fixed Income (Total) index (which has been renamed HFRI Relative Value: Multi-strategy index).

In a more recent study, Duarte, Longstaff, and Yu (2007) created returns series of several fixed income arbitrage trades frequently used by hedge funds—swap spreads, yield-curve spreads, mortgage spreads, fixed income volatility arbitrage, and capital structure arbitrage.

Essentially, the swap spread trade is a bet that the fixed side of the spread (the difference between the swap rate and the yield of the Treasury security of the same maturity) will remain higher than the floating side of the spread (the difference between LIBOR and the repo rate) while staying within a reasonable range that can be estimated from historical data. Yield-curve spread trades are "butterflies", betting that bond prices (which can be mapped to points along the yield curve) deviate from the overall yield curve only for short-run, tactical liquidity reasons, which dissipate over time. Mortgage spread trades are bets on prepayment rates, consisting of a long position on a pool of GNMA mortgages financed using a "dollar roll", delta-hedged with a five-year interest rate swap. Fixed income volatility trades are bets that the implied volatility of interest rate caps tends to be higher than the realized volatility of the Eurodollar futures contract. Capital-structure arbitrage or credit arbitrage trades on mispricing among different securities (for example, debt and equity) issued by the same company.

Duarte, Longstaff, and Yu (2007) found strong correlation between the returns of these strategies and the returns of fixed income arbitrage hedge funds. In addition, many of these strategies have significant exposure to risks in the equity and bond markets. DJCS reports an index of *Convertible Arbitrage Hedge Funds*. A description of the strategies utilized by funds included in this index is as follows:[57]

> The Dow Jones Credit Suisse Convertible Arbitrage Hedge Fund Index[SM] is a subset of the Dow Jones Credit Suisse Hedge Fund Index[SM] that measures the aggregate performance of convertible arbitrage funds. Convertible arbitrage funds typically aim to profit from the purchase of convertible securities and the subsequent shorting of the corresponding stock when there is a pricing error made in the conversion factor of the security. Managers of convertible arbitrage funds typically build long positions of convertible and other equity hybrid securities and then hedge the equity component of

[57] For further information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_CVaRB&cy=USD.

the long securities positions by shorting the underlying stock or options. The number of shares sold short usually reflects a delta neutral or market neutral ratio. As a result, under normal market conditions, the arbitrageur generally expects the combined position to be insensitive to fluctuations in the price of the underlying stock.

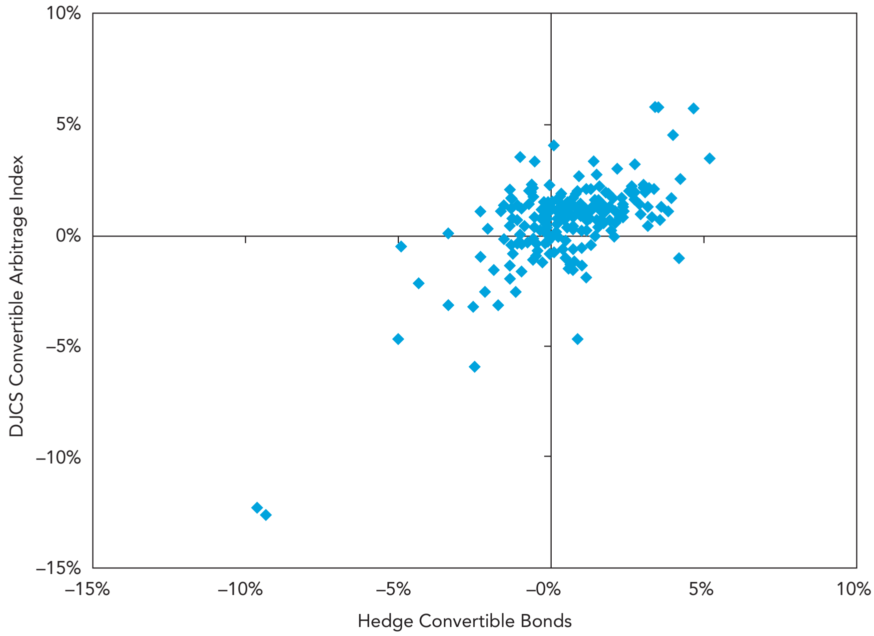


Figure 9.12 **Risk factor for DJCS convertible arbitrage index: 1994–2010.**

Using a sample of US and Japanese convertible bonds, Agarwal et. al. (2010) created the return of a rule-based, passive convertible bond arbitrage strategy which they label as the *"buy-and-hedge"* strategy. The strategy mimics the performance of purchasing a broad portfolio of convertible bonds and mechanically hedges the implicit equity exposure by shorting an appropriate amount of stocks. This strategy resembles the usual passive buy-and-hold strategy of conventional asset-class indices but for the addition of the equity hedge. Figure 9.12 presents a simplified version of the Agarwal et. al. (2010) model by comparing the hedged returns of a broad-based portfolio of convertible bonds to the performance of the DJCS convertible arbitrage index.[58]

The results are consistent with the Agarwal et. al. (2010) findings and confirms. One interpretation of the Agarwal et. al. (2010) results is that the return to convertible arbitrage hedge funds stems from a liquidity premium paid by issuers of convertible bonds to the hedge fund community for holding inventories of convertible bonds, managing the inherent risk by hedging the equity content of these bonds.[59]

DJCS reports an index of *Long/Short Equity Hedge Funds*. A description of the strategy used by these managers is as follows:[60]

> The Dow Jones Credit Suisse Long/Short Equity Hedge Fund Index[SM] is a subset of the Dow Jones Credit Suisse Hedge Fund Index[SM] that measures the aggregate performance of long/short equity funds. Long/short equity funds typically invest in both long and short sides of equity markets, generally focusing on diversifying or hedging across particular sectors, regions or market capitalizations. Managers typically have the flexibility to shift from value to growth; small to medium to large capitalization stocks; and net long to net short. Managers can also trade equity futures and options as well as equity related securities and debt or build portfolios that are more concentrated than traditional long-only equity funds.

This is an important hedge fund style category. The long/short equity style consistently accounts for 30–40% of the total number of hedge funds. Agarwal and Naik (2004) studied a wide range of equity-oriented hedge funds, and Fung and Hsieh (2011) focused on long/short equity funds. The empirical evidence shows that long/short equity funds have directional exposure to the stock market as well as exposure to long small-cap/short large-cap positions, similar to the SMB factor in the Fama and French (1992) three-factor model for stocks.

Figure 9.13 provides support for this view. Here, we use the previous twenty-four months of data to estimate the exposure of long/short equity funds (as proxied by the DJCSI Long/Short Equity Index) to three market factors: the S&P 500 index, the Russell 2000 index, and the MSCI EAFE index. The estimated coefficients are used to perform a one-month-ahead

[58] We used the Vanguard Convertible Securities Portfolio as a proxy for the convertible bond universe. The hedging is done by a rolling regression of the convertible bond portfolio to the Russell 2000 index to estimate the amount of short equity index position needed.

[59] This is analogous to the role played by market makers of securities.

[60] For further information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_LOSHO&cy=USD.

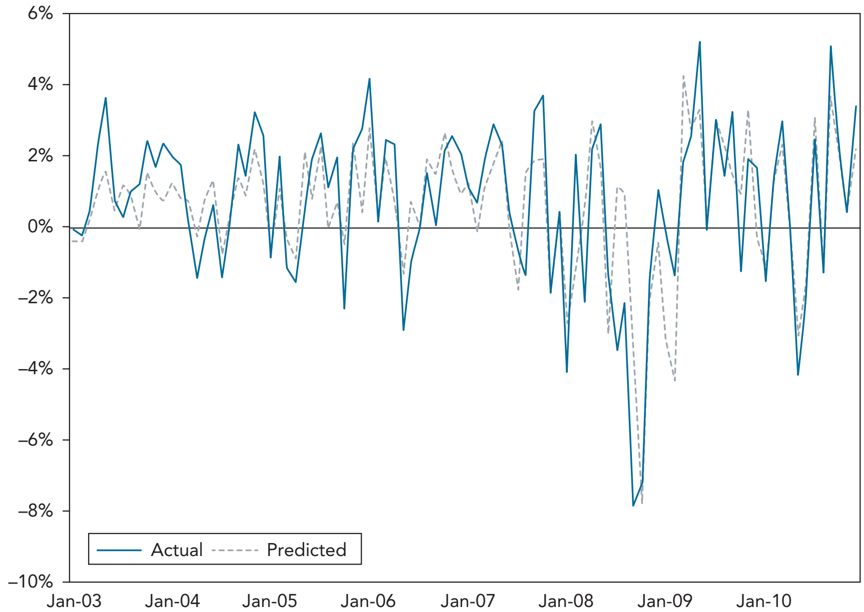


Figure 9.13 **Actual and predicted returns for DJCS long/short equity index.**

Source: Fung and Hsieh (2006) Figure 6 updated.

conditional forecast.[61] The figure shows that the one-month-ahead forecast is a very good predictor of the DJCS Long/Short Equity index. An intuitive explanation of these results is as follows. Typically, long/short equity hedge fund managers are stock pickers with diverse opinions and ability. As such, the individual performance of these managers is likely to be highly idiosyncratic. However, all managers are subject to the basic phenomenon that "underpriced stocks", if they exist, are likely to be found among smaller, "under-researched" stocks, or foreign markets (particularly emerging markets). On the short side, liquidity conditions in the stock-loan market make small stocks and foreign stocks much less attractive candidates for short sales.

Niche Strategies: Dedicated Short Bias, Emerging Market and Equity Market Neutral

The remainder of this section covers the other three DJCS strategy indices. DJCS provides the following description of the Dedicated Short Bias strategy as follows.[62]

> The Dow Jones Credit Suisse Dedicated Short Bias Hedge Fund Index[SM] is a subset of the Dow Jones Credit Suisse Hedge Fund Index[SM] that measures the aggregate performance of dedicated short bias funds. Dedicated short bias funds typically take more short positions than long positions and earn returns by maintaining net short exposure in long and short equities. Detailed individual company research typically forms the core alpha generation driver of dedicated short bias managers, and a focus on companies with weak cash flow generation is common. To affect the short sale, the manager typically borrows the stock from a counterparty and sells it in the market. Short positions are sometimes implemented by selling forward. Risk management often consists of offsetting long positions and stop-loss strategies.

As expected, the Dedicated Short Bias strategy is negatively correlated to equities, which is shown in Figure 9.14. In the regression of the DJCS Dedicated Short Bias index on the SNP index, the slope coefficient is −0.81 (with a *t*-statistic of −16.1) and an R^2 of 0.56.

DJCS provides the following description of the *Emerging Market* strategy as follows:[63]

> The Dow Jones Credit Suisse Emerging Markets Hedge Fund Index[SM] is a subset of the Dow Jones Credit Suisse Hedge Fund Index[SM] that measures the aggregate performance of emerging markets funds. Emerging markets funds typically invest in currencies, debt instruments, equities, and other instruments of countries with "emerging" or developing markets (typically measured by GDP per capita). Such countries are considered to be in a transitional phase between developing and developed status. Examples of emerging markets include China, India, Latin America, much of Southeast Asia, parts of Eastern Europe, and parts of Africa.

[61] Specifically, the one-month-ahead conditional forecasts use the regression coefficients from the previous 24 months and the realized values of the regressors in the subsequent month.

[62] For further information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_DEDSH&cy=USD.

[63] For further information go to http://www.hedgeindex.com/hedgeindex/secure/en/indexperformance.aspx?indexname=HEDG_EMMKT&cy=USD.

Figure 9.14 **Risk factor in DJCS dedicated short bias index: 1994–2010.**

The index has a number of subsectors, including arbitrage, credit and event driven, fixed income bias, and equity bias.

Since it is often very difficult to short securities in the less developed economies, Emerging Market hedge funds typically have a long bias. Figure 9.15 shows clearly that the DJCS Emerging Market index is highly correlated with the MSCI Emerging Market index. The regression of the former on the latter gives a slope coefficient of 0.49 (with a *t*-statistic of 18.6) and an R^2 of 0.63.

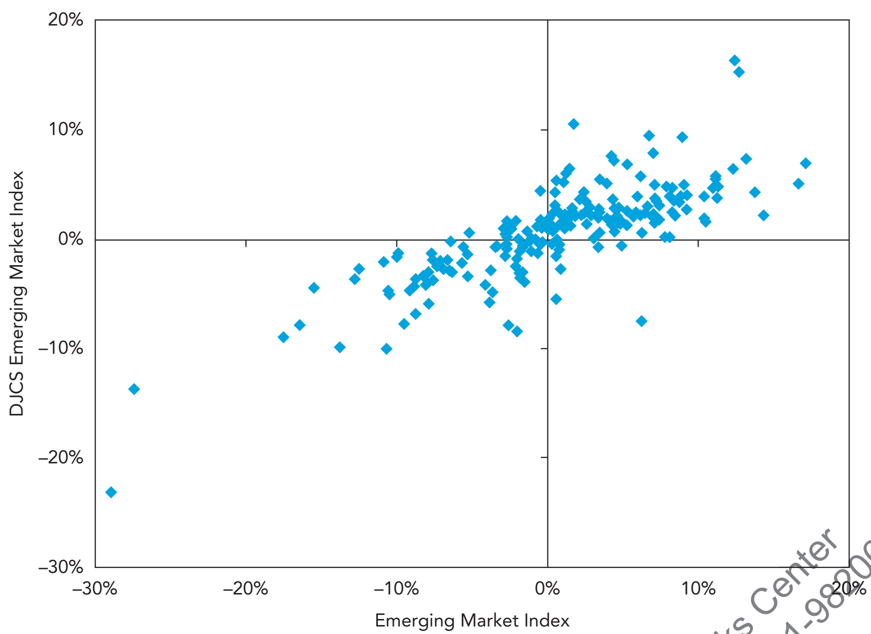


Figure 9.15 **Risk factor in DJCS emerging market index.**

When we examined the collection of hedge funds in the Equity Market Neutral strategy, we did not find a single common component in their returns. This tells us that there is not a single common strategy employed by many funds. Indeed, even index suppliers such as HFR or DJCS differ on which funds are "equity market neutral" funds. Their returns can differ dramatically across different months. It appears that equity market neutral does not behave like a single niche strategy. Return behavior suggests that different funds apply different trading strategies with a similar goal of achieving almost zero beta(s) against a broad set of equity indices. We therefore conclude that there is no single common risk factor that drives the return behavior of Equity Market Neutral funds.

We have shown that, using a bottom-up approach, almost all except one of the DJCS strategy indices can be linked to observable market risk factors. Some of these are standard factors such as equity and bond indices. Others are spread factors, such as the spread between Baa corporate bonds and 10-year treasuries. There are also highly nonlinear factors like volatility factors that behave like portfolios of straddles on bonds, currencies, and commodities. We will discuss the implications of these factors on performance evaluation, portfolio construction and risk management for hedge fund investors.

9.4 WHERE DO INVESTORS GO FROM HERE?

Portfolio Construction and Performance Trend

In the early days of the hedge fund industry, before the academic studies on publicly available hedge fund returns and the attendant risk factors, investors often viewed hedge funds as "black boxes", regarding them as a separate "asset class" that were not correlated to standard equity and bond indices. These early hedge fund investors had limited access to performance data and often had to rely on tools that were designed for evaluating long-bias funds investing predominantly in conventional asset classes. It is perhaps not surprising that portfolio construction and risk management often reduce to spreading risk capital across hedge fund managers with different sounding strategies. During these early days, manager selection was primarily driven by the reputation of individual hedge fund managers. Knowing what we know today, how would we do things differently? Let us go back to the 27 large hedge funds identified in the Fung and Hsieh (2000b) and Fung, Hsieh, and Tsatsaronis (2000). Suppose we are evaluating these managers for potential investments in 1997 at the endpoint of these funds' performance history gathered for the Fung and Hsieh studies. What additional insight can new research undertaken since then add to this decision?

We begin by asking the question: are there systematic risk exposures in their performance? Table 9.3 reports the regression results of these 27 large funds as an equally weighted portfolio (LHF27 for short) against the eight-factor model used to analyze hedge fund indices in Table 9.2. Table 9.3A tells us that over the period 1990–1996 this portfolio has no significant exposure to stocks and bonds.[64] The regression results suggest that returns are partially driven by exposure to directional bets on nonlinear factors in the foreign exchange and commodities markets (PTFSFX-Rf and PTFSCOM-Rf) as well as on emerging markets (IFC-Rf). As an equally weighted portfolio, LHF27 has a highly significant alpha of 1.48% per month. The monthly average total return over this period is 1.58% which implies that total returns are not derived from static exposures to the set of risk factors in the eight-factor model.

Hedge fund managers are primarily active managers, therefore, a static regression model like the eight-factor model used here can only capture the average exposures of these managers' factor bets (betas) over a given sample period. Accordingly we divide this sample period into sub-periods to gain insight on how factor betas may have changed over time.[65] The second and third columns of Table 9.3A report the regression results in two sub-periods, 1990–1993 and 1994–1996 using 1994 as the break point for the analysis of time varying betas. The results are consistent with dynamic adjustments to factor betas responding to changes in the market. Except for a persistent, significant exposure to emerging markets, other factor betas show variability over time. Alpha also appeared to be declining over time (from 2.11% per month during the 1990–1993 period to 0.92% per month during the 1994–1996 period), but remained statistically significant at the 1% level in both sub periods.[66] Overall, total return from the LHF27 portfolio appears to be sensitive to a persistent exposure to the emerging markets and opportunistic bets on a number of other risk factors. It is noteworthy that there is no persistent directional exposure to the US equity market.

How does the performance of the LHF27 portfolio compare to its peers? Unfortunately we do not have reliable history for the DJCS index before 1994. We can however answer this question for the second sub-period 1994–1996. Table 9.3B reports the results of two statedependent regressions of LHF27 to the two indices—DJCSI and HFRI over the period 1994–1996 by adding dummy variables that take on the value of 0 whenever the SNP return is greater than its median return over this sample period and the value of 1 otherwise.

$$\text{LHF27} = \alpha + \text{SNPDUMMY} + \beta^*\text{Index(DJCSI or HFRI)} + \beta 1^*\text{SNPDUMMY}^*\text{Index(DJCSI or HFRI)} + \epsilon \qquad (9.1)$$

This regression equation helps us answer the question of whether there is peer group alpha relative to the average performance of the hedge fund industry (the hedge fund indices) and whether it varies according to the stock market environment (is it a bull or bear market alpha)? The LHF27 portfolio has an alpha of 1.89% (1.65%) per month and a beta of 0.47 (0.21) versus the DJCSI (HFRI) index; all coefficients are highly statistically significant during months when the SNP index return is above its median. The corresponding figures for months when the SNP index return is below it median are an alpha of 1.37% (1.48%)

[64] We chose 1990 as the starting point as some of the instruments used in constructing our dataset do not extend back beyond 1990.

[65] Ideally one should explicitly model the time-varying behavior of factor-betas. However data limitations present serious challenges—see for example Bollen and Whaley (2009) and Patton and Ramadorai (2012).

[66] The null hypothesis of equality of individual coefficients between the 1990–1993 and 1994–1996 sub-periods can be rejected at the 1% significance level for the constant term and the bond straddle (PTFSFX-Rf). In addition, the joint test for the equality of all coefficients is rejected as well.

Table 9.3 Regression of 27 Large Hedge Funds (LHF27), the DJCS and HFRI Indices on an Eight-Factor Model Similar to That of Table 9.2: SP-Rf is the Excess Return of the S&P 500 Index. RL-SP is the Return of the Russell 2000 Index Minus the Return of the S&P500 Index. TY-Rf is the Excess Return of US Ten-Year Treasuries. BAA-TY is the Return of Moody's BAA Corporate Bonds Minus the Return of US Ten-Year Treasuries. PTFSBD-Rf is the Excess Return of a Portfolio of Bond Straddles. PTFSFX-Rf is the Excess Return of a Portfolio of FX Straddles. PTFSCOM-Rf is the Excess Return of a Portfolio of Commodity Straddles. IFCRf is the Excess Return of the International Finance Corporation's Emerging Market Index. Regression Equation (1): LHF*27* = constant + SNPDUMMY + β*Index + β*SNPDUMMY*Index + ϵ

3A.	**LHF27**		
Sample Period	*1990–1996*	*1990–1993*	*1994–1996*
Constant	0.0148	0.0211	*0.0092*
	0.0020	*0.0016*	*0.0028*
	7.4949	**13.2255**	**3.2809**
SP-Rf	0.0606	0.0411	0.2347
	0.0770	*0.1072*	*0.1268*
	0.7869	0.3831	1.8508
RL-SP	0.0701	0.1299	0.0512
	0.0573	*0.0835*	*0.0851*
	1.2241	1.5556	0.6013
TY-Rf	0.2733	−0.1735	0.2069
	0.1502	*0.1707*	*0.1212*
	1.8202	−1.0163	1.7066
BAA-TY	−0.1624	−1.0077	0.3309
	0.2246	*0.3006*	*0.3111*
	−0.7228	**−3.3527**	1.0637
PTFSBD-Rf	0.0272	0.0547	−0.0014
	0.0180	*0.0105*	*0.0179*
	1.5134	**5.1856**	−0.0763
PTFSFX-Rf	0.0236	0.0119	0.0291
	0.0088	*0.0130*	*0.0105*
	2.6911	0.9151	**2.7750**
PTFSCOM-Rf	0.0499	0.0265	0.0522
	0.0146	*0.0167*	*0.0256*
	3.4235	1.5831	**2.0357**
IFC-Rf	0.1171	0.1268	0.1412
	0.0281	*0.0396*	*0.0605*
	4.1757	**3.2067**	**2.3332**
Adj R^2	0.3345	0.3620	0.4303
D.W.	1.7778	1.8685	2.1050

Source: Fung and Hsieh, 2000b.

Table 9.3 Continued

3B. 1994–1996	Constant	SNPDUMMY	Index	SNPDUMMY* Index	Adj Rsq	Durbin Watson	*p*-Value of Test of Equality	
DJCSI	0.0188	−0.0137	0.4657	0.3018	0.3251	1.6430	Constant	0.0000
	0.0026	0.0034	0.1249	0.1338			Slope	0.0241
	7.2308	**−4.0294**	**3.7286**	**2.2556**			Joint	0.0000
HFRI	0.0165	−0.0148	0.2065	0.8193	0.1811	1.8115	Constant	0.0099
	0.0045	*0.0057*	*0.2297*	*0.2681*			Slope	0.0022
	3.6667	**−2.5965**	0.8990	**3.0559**			Joint	0.0031

SNPDUMMY = 1 (SP index return < = median); Newey-West (1987) standard errors with 6 lags in italics.

and a beta of 0.30 (0.82) versus the DJCSI (HFRI) index; all coefficients are highly statistically significant. The joint test of equality indicates that there is a nonlinear correlation between these large funds and the indices. Taken together these results are consistent with a superior performing portfolio, LHF27, relative to the market averages (hedge fund indices) achieved by dynamically managing it risk exposure (betas) in response to changing equity market conditions. This is consistent with the eight-factor model results, which tell us that these large hedge funds take time-varying bets on different risk factors, and their overall risk profile relative to their peers (hedge fund index betas) is nonlinear and responsive to changing equity market conditions.

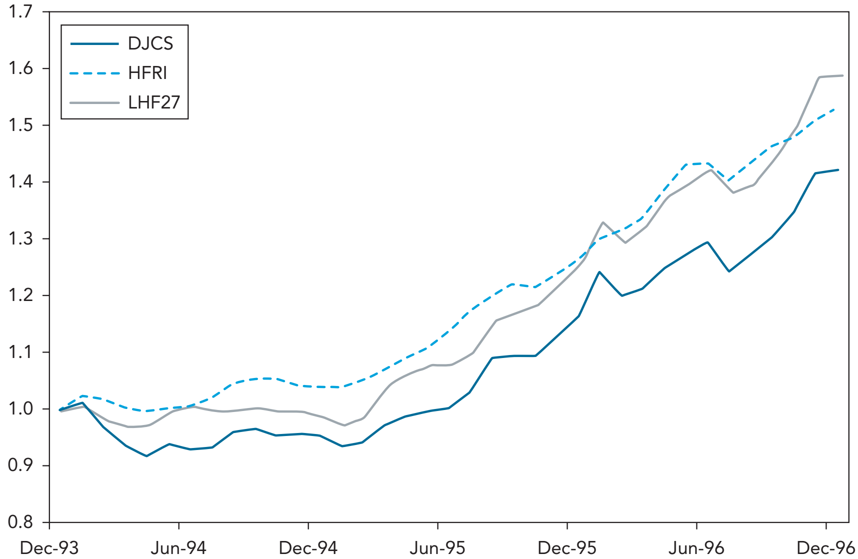


Figure 9.16 Cumulative return of DJCSI, HFRI indices and the LHF27 portfolio: 1994–1996.

Figure 9.16 plots the cumulative performance of the LHF27, DJCSI, and HFRI. It shows that the LHF27 portfolio delivered a cumulative return of 59.08% versus 42.24% and 53.18% respectively for the DJCSI and HFRI index. The respective annualized returns for the period (1994–1996) are 15.88%, 12.20%, 14.42% for LHF27, DJCSI, and HFRI with corresponding annualized standard deviations of 8.04%, 9.08%, and 4.94%. As we pointed out, data prior to 1996 from commercial databases are susceptive to backfilled and survivorship biases, whereas the LHF27 portfolio suffers from ex post selection bias. Just how much will measurement biases affect these impressive historical performance statistics?

How Much of the LHF27 Portfolio's Monthly Alpha of 2.11% (1990–1993) and 0.92% (1994–1996) Is Due to Measurement Bias?

In the absence of a commercially available hedge fund performance database, this would be a very difficult question to answer in 1997. Let us see what hindsight would have told us about two important biases that overshadow the reliability of the LHF27's historical alpha. Unfortunately the available data over the 1997 to 2001 period is insufficient to support definitive answers to this question. Given the hedge fund industry's change in investor clientele post 2001, it is instructive to

Table 9.4A Assets Under Management of the Top 50 Reporting Funds to Commercial Databases ($ billion): At the End of Each Formation Year, Funds Reporting to Three Commercial Databases (BarclayHedge, HFR, Lipper-TASS) are Sorted According to Their Assets Under Management (AUM). Duplicated Funds are Eliminated. Top 50 is the Sum of the AUM of the Top 50 Funds in This Ranking. All Other Funds Denotes the Sum of AUM of All Other Funds Below the Top 50

	Formation Year									
AUM in $ billions	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
All other funds	132	168	231	350	456	570	842	539	541	666
Top 50	70	94	113	133	166	209	265	233	228	278
Total AUM	202	262	343	483	622	779	1,107	772	769	945
Top 50/Total AUM (%)	34.65	35.88	32.94	27.54	26.69	26.83	23.94	30.18	29.65	29.42

Source: BarclayHedge, HFR, Lipper-TASS.

Table 9.4B AUM Range of Top 50 Reporting Funds to Commercial Databases ($ Million) Versus Overall Median Fund Size. At the End of Each Calendar Year Funds are Ranked by AUM to Identify the Top 50 Hedge Funds

	Formation Year									
Break Points in $ millions	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Overall median fund size	25	29	30	33	33	35	43	23	28	23
Top 50—smallest	693	806	1,066	1,476	1,748	2,300	3,042	2,456	2,381	2,638
Top 50—largest	7,878	8,557	9,534	7,793	16,580	16,045	11,685	13,986	12,500	22,662
Smallest/Median	27.72	27.79	35.53	44.73	52.97	65.71	70.74	106.78	85.04	114.70
Largest/Median	315.12	295.07	317.80	236.15	502.42	458.43	271.74	608.09	446.43	985.30

Source: BarclayHedge, HFR, Lipper-TASS.

examine this performance persistency question using a dataset that spans the recent decade. Table 9.4 tracks the Top 50 funds ranked by AUM in each of the sample year based on end of year figures reported to the three consolidated commercial databases BarclayHedge, HFR, Lipper-Tass.

At the end of each year, we identify a set of unique funds that report to the three commercial databases by eliminating all duplicated funds whereby creating a consolidated database almost free of doubted counted funds. We then sort all these hedge funds in the consolidated database by AUM and identify the Top 50 funds. Dividing the total sample into two sets of funds—the Top 50 and all others—we record the total year end AUM for each set and tabulate them in Table 9.4A. Table 9.4B reports the median fund's year end AUM and the AUM range (smallest and largest) of the Top 50 subset for each sample year 2001 to 2010. The results show that the Top 50 funds consistently manage between 23.94% to 35.88% of the hedge fund industry's total AUM. Table 9.4B shows that the Top 50 funds are many times bigger than the median fund in our consolidated database—ranging from 27.72 times (smallest Top 50 in 2001) to 985.30 times (largest Top 50 in 2010). This size divergence between the largest fund and the median fund has been steadily growing in this first decade of the millennium.

Table 9.4C tabulates the rate of entry and exit of funds in our consolidated database. When a fund exits commercial databases, there are typically two broad categories of reasons. One, it is being dropped by the database vendor because it no longer meets the vendor's inclusion criteria. Two, the fund manager elects to stop reporting. In the first category, database vendors attempt to flag those exiting funds that went out of business but the information is generally incomplete. In the second category, the reasons can be many and varied; ranging from a dying fund whose manager no longer finds reporting bad news to a database to be a business priority, to successful funds who have reached their AUM capacity limit and find the burden of reporting performance to the generally public a costly exercise with

Table 9.4C **Entry, Exit, and Survival Rates of Top 50 and All Other Funds. At the End of Each Calendar Year, We Compute the Fraction of New Funds Entering Into, Surviving Funds From Last Year, Liquidated and Unknown Exits from Our Combined Database Relative to the Total Number of Unique Funds in Our Sample**

		Formation Year								
		2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
All others	New Entry in Formation Year	0.07	0.08	0.08	0.11	0.10	0.09	0.08	0.06	0.06
	Survive Next Year	0.86	0.85	0.87	0.85	0.83	0.81	0.72	0.76	0.82
	Exit Next Year/Liquidated	0.08	0.08	0.07	0.10	0.10	0.09	0.14	0.11	0.01
	Exit Next Year/Unknown	0.06	0.07	0.05	0.06	0.07	0.11	0.15	0.13	0.17
	Total Exit	0.14	0.15	0.13	0.15	0.17	0.20	0.29	0.24	0.18
Top 50	New Entry in Formation Year	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
	Survive Next Year	0.98	0.92	0.94	0.98	0.88	0.84	0.82	0.94	0.88
	Exit Next Year/Liquidated	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00
	Exit Next Year/Unknown	0.02	0.08	0.06	0.02	0.12	0.16	0.14	0.06	0.12
	Total Exit	0.02	0.08	0.06	0.02	0.12	0.16	0.18	0.06	0.12

diminishing benefits.[67] Suffice to say that large hedge funds are likely to fall into this second category. Therefore, when a large fund drops out of a database, it does not necessarily carry the negative interpretation of the usual empirical studies on survivorship bias. In fact, the bias can easily go the other way. The results in Table 9.4C illustrate this point. Comparing the two "New Entry to Database in Formation Year" rows, we see the rate at which new funds enter the database for all other funds whereas only once in 2007 we had a new Top50 fund that first reports performance to the databases that year.[68] Comparing the two "Exit Next Year/Liquidated" rows, we see that only once we saw eight funds (4%) of the 2007 Top 50 funds met with liquidation the following year (2008). In all other cases, exits from the Top 50 subset tend to be selfselected for reasons unknown to the database vendors. The evidence suggests that large funds are better at retaining investors' assets and staying in business than smaller funds.[69] However, there still exists the possibility that a Top 50 fund in any given year can get demoted when its AUM falls below the threshold for that year. It is not always clear that demoted funds necessarily underperform promoted funds that took their places from year to year. We provide some insight to this question later. For now, the empirical evidence thus far, when taken together, points to a smaller measurement bias in the estimated 1.48% monthly alpha of the 27 large-hedge fund portfolio than that of the average hedge fund. If these impressive high-teen returns with limited persistent exposure to systemic risks (factor betas) cannot be easily accounted for by measurement bias it begs the question "is the attractive performance of large hedge funds a sustainable proposition?"

Simulating the Performance of Investing in Large Funds

To gain insight into the strategy of investing into a portfolio of large hedge funds, we synthesize a comparable experiment to the LHF27 portfolio by constructing an equally weighted portfolio of the largest 50 hedge funds (TOP50 for short) using more recent data—2002–2010. We do this in two ways. First, to minimize ex post selection bias, we construct a TOP50 portfolio as follows. Starting at the end of 2001, take the Top 50 funds ranked by 2001 year end AUM to create an equally weighted portfolio (TOP50) of these 50 funds and observe their returns for the following year.[70] The process is repeated at the end of each

[67] There are other reasons such as those funds that change the vendors they report to etc. See Fung and Hsieh (2009) for a discussion on non-reporting funds.

[68] The entry to Table 9.4C shows a 2% figure which is one fund for the Top 50 subset.

[69] This is consistent with the conclusion reached in Fung and Hsieh (2009) and some recent results in an unpublished paper in Edelman, Fung and Hsieh (2011). Because large funds tend to stay large, the lookahead bias of selecting large funds ex-post is also less severe. However we caution the reader that our conjecture is derived from available data which remains incomplete to-date.

[70] In the event of one of these Top 50 funds exiting the databases, its capital is reinvested in cash to be conservative in our simulation.

year until the end of 2009, resulting in a return series mimicking the performance of a strategy of investing in the Top 50 hedge funds in equal dollar amounts while rebalancing at the end of each calendar year resulting in a return series spanning 2002–2010. Second, we create a TOP50_2010 portfolio by computing the pro forma returns of the Top 50 hedge funds created at the end of the sample period based on 2010 year end AUM. This second portfolio has a similar ex-post selection bias to the LHF27 portfolio, in that the identities of the top 50 large funds at the end of the sample period is not known during the earlier years. Comparing the TOP50 and TOP50_2010 portfolios gives us a lower and upper bound of the performance investors can achieve by simply adopting the strategy of investing equally into the Top 50 large funds and rebalancing each year (lower bound) and that of a foresight-assisted portfolio TOP50_2010 (upper bound).

Table 9.5 reports the performance characteristics of the TOP50, DJCSI, and HFRI versus the eight-factor model in the same format as Table 9.3. Table 9.5A shows the TOP50 portfolio to have no statistically significant alpha whereas the foresight-assisted portfolio TOP50_2010 shows a monthly alpha of 0.53% and is statistically significant at the one percent level. This is a significant decline from the results we saw with the LHF27 portfolio. However, it must be noted that there is no significant negative alpha. This is important as the returns we used for the TOP50 portfolio is net of all fees and expenses whereas no trading cost is factored into the returns of the risk factors in the regression model. This micro effect of trading friction will be discussed further in a later section.[71] The decline in monthly alpha from 2.11% (1990–1994) to 0.92% (1994–1996) for the LHF27 portfolio and now to 0% and 0.52% respectively for the TOP50 and TOP50_2010 portfolios over the recent decade (2002–2010) is consistent with the combined effect of rising competition in a hedge fund industry that experienced almost a ten-fold increase in AUM since 1996—an outcome that is anticipated by the theoretical model of Berk and Green (2004). In terms of risk factor exposures, like the LHF27, the TOP50 and the TOP50_2010 portfolios continue to exhibit highly significant betas to the emerging market factor (IFC-Rf). While both portfolios have no significant beta to the US equity factors, they both have a highly significant credit-factor (BAA-TY) beta.

Relative to their peers, Table 9.5B shows that both the TOP50 and the TOP50_2010 portfolios have statistically significant alpha relative to the hedge fund indices—DJCSI and HFRI. The SNPDUMMY variable shows a significant lowering of the DJCSI alpha during months when the SNP index return is below its median. Overall, the results indicate that the strategy of buying large hedge funds delivers superior performance to investing in indices (or the industry average). In terms of dynamic risk-taking behavior the results hint at a nonlinear, stock market dependent shift for the TOP50_2010 portfolio. This is a subject worthy of further investigation as we accumulate more data after the 2008 financial crisis.[72]

Cumulatively, over this 2002–2010 period, all of the portfolios TOP50, DJCSI and HFRI outperformed the equity market (proxy by the SNP index) as Figure 9.17 shows.[73] These results taken together are supportive of the continuing interest of institutional investors in hedge funds; and especially in large hedge funds. Capital continues to flow into the hedge fund industry after the 2008 financial crisis. The total industry AUM figures in Table 9.5A show the 2010 AUM figure is once again on the rise after a pause following the 2008 crisis. Although the Top 50 hedge funds' total AUM remains around 30% of total industry AUM, the size of these larger hedge funds relative to the median of our data universe continues to expand in favor of larger funds. Our analysis of large hedge funds shows performance statistics consistent with managers of these funds continuing to deliver alpha relative to their peers and having low exposure to the US equity market which helps to explain their success in attracting institutional investors' capital. Later, we discuss other market structural issues that reinforces the persistency of this capital formation trend.

9.5 RISK MANAGEMENT AND A TALE OF TWO RISKS

An important feature that attracts investors to the hedge fund industry is the variety of strategies hedge fund managers deploy and the diversity of assets to which these strategies are applied. Even at an aggregated level, long-term historical return patterns reveal several categories of trading styles with low return correlations. Commercial index providers like DJCS reports ten distinct style-specific sub-indices.[74] Under normal circumstances,

[71] A similar finding was reported in Fung et. al. (2008) and Edelman et. al. (2012) who also reported a similar trend decline of historical alpha.

[72] We also ran a similar regression to Equation (9.1) using the credit spread variable (BAA-TY) as the conditioning risk factor instead of the SNP. The results are similar.

[73] The cumulative returns are 76.62%, 91.55%, 86.70%, and 30.59% respectively for TOP50, DJCS, HFRI and SNP.

[74] See http://www.hedgeindex.com/hedgeindex/en/indexoverview.aspx?indexname=HEDG&cy=USD for more details, excluded are those sub-indices that are combinations of other specialist styles.

Table 9.5 In Panel A Regression of TOP50 and TOP50_10 Indices on Eight-Factor Model Similar to that of Table 9.2: SP-Rf is the Excess Return of the S&P 500 Index. RL-SP is the Return of the Russell 2000 Index Minus the Return of the S&P 500 Index. TY-Rf is the Excess Return of US Ten-Year Treasuries. BAA-TY is the Return of Moody's BAA Corporate Bonds Minus the Return of US Ten-Year Treasuries. PTFSBD-Rf is the Excess Return of a Portfolio of Bond Straddles. PTFSFX-Rf is the Excess Return of a Portfolio of FX Straddles. PTFSCOM-Rf is the Excess Return of a Portfolio of Commodity Straddles. IFC-Rf is the Excess Return of the International Finance Corporation's Emerging Market Index, and in Panel B We Run the Regression Equation (1) Using: TOP50(TOP50_2010) = constant + SNPDUMMY + β*Index + β*SNPDUMMY*Index + ϵ

5A.	TOP50	TOP50_10 Pro Forma
Sample Period	*2002–2010*	*2002–2010*
Constant	0.0007	0.0053
	0.0011	*0.0009*
	0.6291	**5.8624**
SP-Rf	0.0170	0.0071
	0.0342	*0.0431*
	0.4981	0.1655
RL-SP	0.0176	0.0258
	0.0456	*0.0458*
	0.3863	0.5631
TY-Rf	−0.0312	0.0199
	0.0564	*0.0627*
	−0.5526	0.3169
BAA-TY	0.1286	0.1107
	0.0466	*0.0337*
	2.7624	**3.2831**
PTFSBD-Rf	−0.0165	−0.0073
	0.0093	*0.0090*
	−1.7864	−0.8097
PTFSFX-Rf	0.0079	0.0129
	0.0034	*0.0050*
	2.3325	**2.5533**
PTFSCOM-Rf	0.0048	0.0019
	0.0074	*0.0058*
	0.6466	0.3284
IFC-Rf	0.1091	0.1377
	0.0199	*0.0268*
	5.4711	**5.1310**
Adj R^2	0.5402	0.5421
D.W.	1.9186	1.9674

Newey-West (1987) standard errors with 6 lags in italics.

5B.	TOP50		TOP50_2010	
Index = Sample Period	*HFRI*	*DJCS*	*HFRI*	*DJCS*
	2002–2010		**2002–2010**	
Constant	0.0024	0.0031	0.0063	0.0078
	0.0009	*0.0011*	*0.0012*	*0.0012*
	2.6667	**2.8182**	**5.2500**	**6.5000**
DUMSP	−0.0024	−0.0029	−0.0021	−0.0037
	0.0015	*0.0016*	*0.0015*	*0.0017*
	−1.6000	−1.8125	−1.4000	**−2.1765**
Index	0.544	0.4207	0.6281	0.4484
	0.0496	*0.0470*	*0.0752*	*0.0657*
	9.9677	**8.9511**	**8.3524**	**6.8250**
Index * DUMSP	−0.0331	−0.0315	−0.1478	−0.099
	0.0680	*0.0539*	*0.0789*	*0.0598*
	−0.4868	−0.5844	−1.8733	−1.6555
Adj R^2	0.7053	0.6952	0.7162	0.6531
D.W.	1.6547	1.6052	1.687	1.6323
p-value of test of equality				
Constant	0.1048	0.0625	0.1663	0.0302
Slope	0.6261	0.5588	0.0609	0.0981
Joint Test	0.2351	0.0862	0.0352	0.0087

SNPDUMMY = 1 (SP index return ≤ median); Newey and West (1987) standard errors with 6 lags in italics.

these sub-indices perform quite independently of each other as some of the risk factor analysis shows. However, periodically, market events occur during which seemingly different strategies can undergo stress at the same time. When this occurs, an otherwise diverse portfolio of hedge funds can converge in terms of risk, or its portfolio diversification implodes. To gain insight on how convergence of risk factors can be managed we analyze some of these incidents to underscore the special nature of hedge fund strategies.

The first such recorded event occurred during the March to April period in 1994. Seven out of the ten style-specific sub-indices in the DJCS family lost money in these two months. The exceptions being Short Sellers and Managed Futures which delivered positive performance in both months, Risk Arbitrage with a positive return in March and Equity Market Neutral with a positive return in April. The overall DJCS Broad index is negative for both these two months. The reason for the negative performance across strategies is generally attributed to liquidation of leveraged positions following the unexpected rate hike by the US Federal Reserve. The next episode occurred in August 1998 leading up to the collapse of LTCM. In this month, eight out of the ten niche DJCS style sub-indices suffered sizeable losses ranging from −23.03% for emerging market hedge funds to −0.85% for equity market neutral hedge funds. The exception being Short Sellers and Managed Futures funds which returned 22.71% and 9.95% respectively. Much has been written about the events leading up to the collapse of LTCM,[75] suffice it to say that the commonly accepted causes are market-wide liquidation of risky assets and the bloated balance sheet of LTCM from aggressive use of leverage. While wholesale liquidation of risky assets can occur globally for a variety of reasons inflicting stress on conventional long-biased strategies as well as hedge funds, it is the effect of leverage that makes hedge fund investment risk different from conventional long-bias strategies. Put differently, managing hedge fund investment risk applies to both sides of the

[75] See for example Lowenstein (2001).

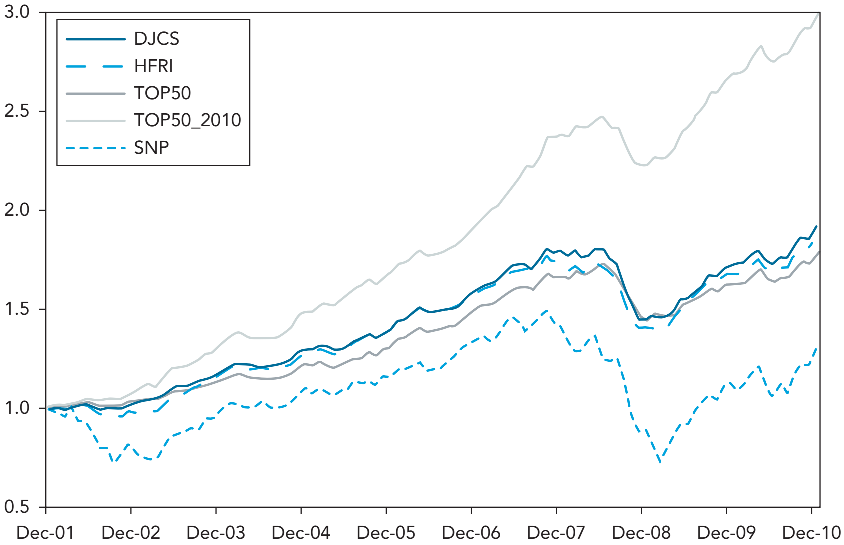


Figure 9.17 Cumulative return of DJCS, HFRI, SNP indices, and the TOP50 and TOP50_2010 portfolios: 2002–2010.

balance sheet—asset as well as liability—in that a diversified set of risk factor exposures from the asset side of the balance sheet is no guarantee that funding risk is simultaneously diversified.

Events leading up to the 2008 financial crisis are a stark illustration of the damage that a market-wide funding crisis can inflict on leveraged positions. August 2007 marked the first ever month in the reported history of DJCS's hedge fund indices in which all nine specialist style sub-indices lost money—the only exception being the positive return from specialist funds with a short focus or short sellers. Some authors have attributed the event to "firesale liquidation of similar portfolios that happened to be quantitatively constructed."[76] This, however, would not explain how all other hedge fund styles lost money as well. Another clue lies with the extraordinary events leading up to the demise of Bear Stearns.[77] Bear Stearns was among the first major Wall-Street victims of the credit contagion emanating from what we know with hindsight as the US housing market crisis. Around the peak period of the 2008 financial crisis between July and October of 2008, there were unprecedented consecutive months, July to September, during which all hedge fund styles except short sellers lost money. The precise reasons why different specialist hedge fund styles lost money are likely to be many and varied. The one common theme is the "forced liquidation" of leveraged positions driven by a combination of rising margin requirements, borrowing costs and investors withdrawing their equity. In short, there was a crisis on the liability side of most hedge funds' balance sheet irrespective of the type of assets (risk factors) they hold (are exposed to) and their trading strategy. No major strategy category was spared. For example, a historically low volatility strategy like Convertible Arbitrage lost over 12% a month for both September and October of 2008 according to the DJCS style index for that strategy.[78] These were the two left tail data points of this strategy shown in Figure 9.12. Under normal market conditions, shorting the equity content of a convertible bond will eliminate most of the day-to-day price fluctuation thus creating a low risk portfolio to take advantage of perceived mispricing of convertible bonds. However, when these positions are leveraged, the latent risk of rolling over these positions can prove to be extremely costly in a credit crunch.

Can funding risk be mitigated in part or in whole? It is now almost four years since the 2008 financial crisis and the impact of global deleveraging is still unfolding with no end in sight. The profitability of most hedge fund strategies is driven by effective use of leverage. Therefore the risk of another funding crisis similar in character to what we experienced in the summer of 2008 is something that cannot, and should not, be overlooked. It is also a risk that cannot be easily mitigated by simply spreading one's capital to different hedge fund strategies. The need to consider hedging a credit-driven tail risk event in a hedge fund portfolio is clear; however, a complete solution to this problem remains beyond our grasp and lies beyond the scope of this chapter. Nonetheless, we offer the readers a few "empirical fruits" for thought. Managed futures, as a hedge fund style, has historically been a strategy that delivers a convex performance profile relative to other hedge fund strategies.[79] Unfortunately, on the whole, the DJCS Managed Futures index failed to deliver mitigating performance during the losing months of August and September of 2008. However the replication of trend following as a sub category of this strategy group could be an interesting substitute to the DJCS Managed

[76] See Khandani and Lo (2007, 2011).

[77] See Hajim and Lashinsky (2007).

[78] See Palazzolo (2009) for more detailed discussions.

[79] See for example Figure 9.6, and for more discussions on this see Fung and Hsieh (2004b).

Futures index.[80] The subject of replicating and separating risk factors inherent in risk hedge fund strategy will be taken up further. Suffice to say that the risk management of hedge fund portfolios must take into consideration nonlinear risks such as rare but dramatic market events the manifestation of which can take place on either the asset side and/or the liability side of a hedge fund's balance sheet.

9.6 ALPHA-BETA SEPARATION, REPLICATION PRODUCTS, AND FEES

Much has been written about the generous fee structure which hedge funds command. After two decades of nearly unabated growth, the hedge fund industry is much larger (in AUM terms) and dominated by sophisticated institutional investors—investors who are not known for their generosity when it comes to fees. Yet, there is little sign that hedge fund fees are falling at a pace comparable to the industry's asset growth. In this section we discuss some of the perceived value of investing in hedge fund strategies, and how the supply side of hedge fund products has responded and evolved to changing investors' demands. The earlier sections of this chapter laid out a framework for analyzing hedge fund performance and concluded that hedge fund returns come packaged with an alpha-like component mixed with systemic risk exposures (betas). Some of these betas can be traced to asset-class factors that can be accessed via lower cost vehicles.[81] For example Table 9.2 shows that the eight-factor model explains 78.38% of monthly return variation of a hedge fund index like the HFRI. To the extent that the return from a persistent exposures to a factor beta represents the risk premium earned from placing capital at risk, it raises the question: *can these risk premia from different factors be captured via lower cost alternatives to hedge fund vehicles?*

Let us take a bottom-up approach to analyze this question. It shows that some of the more mature hedge fund strategies can be replicated with rule-based models using liquid assets that are readily executable in the market place. Therefore, a rule-based representation of a given hedge fund strategy can be thought of as the beta factor of that strategy or its strategy beta. Collectively, these strategies are often referred to as alternative betas. Towards the second half of 2007 products attempting to replicate hedge fund returns began to emerge. After several years of development, these hedge fund (strategy) beta products are becoming ubiquitous. For example, Credit Suisse overviews their products in this space, Liquid Alternative Betas (LAB), as follows:

> Liquid Alternative Beta (LAB) is a series of indices that aim to replicate the aggregate return profiles of alternative investments strategies using liquid, tradable instruments. LAB benefits directly from Credit Suisse's experience in the alternative investment space, and is supported by a comprehensive quantitative platform with a focus on delivering alternative returns with liquidity and transparency.
>
> LAB strategies utilize quantitative processes in order to provide similar risk/return characteristics to alternative indices and strategies by identifying and sizing exposure to proxy factors and/or systematically implementing trades typically used by alternative investment managers.
>
> *Source:* http://alternativebeta.credit-suisse.com/altbeta/en/ab_home.aspx?cy=USD.

Today, Credit Suisse offers a suite of these LAB products mimicking the factor-related returns of the DJCSI as well as sub-indices like Event-Driven, Long/Short Equity, Risk Arbitrage, and Managed Futures. These products typically offer liquidity terms comparable to mutual funds and charge much lower fees than hedge funds. Other variations to the Credit Suisse's LAB that mimic a broad-based index like the DJCSI are widely available—see for example the Goldman Sachs Absolute Return Tracker (ART) Fund offer in a US mutual fund format, Barclays Hedge Fund Replicator Fund offer in a European (UCITS) mutual fund format and ProShares offer HDG Hedge Replication, an ETF based on the Merrill Lynch Factor Model of hedge fund betas.[82] It is too early to conclude whether these products offer a more cost-effective way to access the systematic component of hedge fund strategies' performance. However, their existence offers a way of separating the alpha component of a hedge fund portfolio's return from its betas in a practical tradable form. The debut of these products also adds to the risk management tool kit available to hedge fund investors. Figure 9.3 shows the convex return characteristics of managed futures strategies (or trend followers). Hedge fund investors concerned with their portfolio's tail risk exposure can modify the return convexity of their portfolio via these liquid, low cost products that have the flexibility of a publicly traded fund like ETFs.[83] If alpha and beta can be

[80] See for example http://faculty.fuqua.duke.edu/~dah7/HFRFData.htm; and http://faculty.fuqua.duke.edu/%7Edah7/DataLibrary/TF-Fac.xls which show very large positive returns in August and September of 2008.

[81] This is not to say that hedge fund managers physically transact in these factors, but that combinations of these risk factors can closely mimic the performance of a variety of matured hedge fund strategies.

[82] See http://www.goldmansachs.com/gsam/advisors/products/featured_funds/art-fund/index.html and http://www.proshares.com/funds/hdg.html.

[83] Recall that most hedge fund vehicles require cumbersome redemption notice and only allow periodic subscriptions, whereas these alternative products are generally deals at daily Net Asset Value (NAV).

satisfactorily separated, it begs the question: *what are the implications on fees levied by conventional hedge fund vehicles?*

To gain insight on the historical trend on the cost of hedge fund investing, we begin by observing the experience of the early institutional investors in hedge funds at the turn of the century. As noted, most of these early institutional investors outsourced the lion's share of the portfolio management function to funds-of-hedge funds. FOHFs offer a one-stop vehicle that delivers manager selection, due diligence, portfolio construction, administration, reporting, and risk management of a diverse group of hedge fund managers located in different parts of the world engaging in leveraged, dynamic strategies transaction in financial centers globally. Prima facie this appears to be a reasonable step to take to access a then new, unfamiliar class of investment strategies. As is always the case, the value of such a service depends on a comparison of the lower net return to an investor versus the cost of engaging the services of FOHFs.

We turn now to multi-manager hedge fund products. Fung et. al. (2008) found that only a small fraction of FOHFs delivered positive and statistically significant alpha above the seven-factor model of Fung and Hsieh (2004a). Interestingly, though, almost no FOHFs delivered negative and statistically significant alpha. Basically, the majority of FOHFs had no alpha, which is not altogether a bad outcome when compared with mutual funds, the vast majority of which produced negative alpha relative to their respective benchmarks. Yet, the observation that few FOHFs delivered alpha means that the fees charged by the hedge funds and the FOHFs accounted for all excess returns (or alpha) generated before fees. Nonetheless, the lack of persistent negative alpha does imply that hedge fund managers do generate alphas before fees which vanished after adjusting for fees; a phenomenon consistent with the prediction of Berk and Green (2004)'s theoretical model. However, in a general equilibrium context, new hedge fund products will emerge to offer lower fee hedge fund vehicles. It is perhaps not surprising that around 2003, passive, rule-based multi-manager hedge fund products began to emerge competing for investors' capital. Today, Credit Suisse offers a suite of multi-manager portfolios (All Hedge Indexes) that mimics the family of DJCS hedge fund indices—see http://www.hedgeindex.com/hedgeindex/en/indexoverview.aspx?indexname=SECT&cy=USD. Hedge Fund Research (HFR) offers a number of similar multi-manager portfolios, see https://www.hedgefundresearch.com/hfrx_reg/index.php?fuse=login_bd. Since their inception, the tracking errors of these passive investable hedge fund indices have been consistently positive, in the sense that the target index consistently outperformed the tracker portfolio. Actual operational costs of these portfolios alone cannot completely explain the sizeable tracking errors, which average over 2% per annum since their inception depending on the actual target tracked. We believe two major factors contribute to these tracking errors. First, investable indices offer better liquidity terms than the typical hedge fund. Therefore one would expect returns from investable indices to underperform their target by a liquidity premium. Second, hedge fund indices are statistical constructs and as such, they carry index construction rules that cannot be implemented in practice. For example, the HFRI index assumes rebalancing at the end of each month to maintain equal dollar invested in each of its index components. This is clearly not possible as nearly all hedge funds require advance notification for exiting investors ranging from 30 days to 90 days, and in some cases, redemptions are only permitted on an annual basis while others carry an early exit penalty—see Fung and Hsieh (2004b) for more discussions on this topic.[84]

We now return to our simulation of the large hedge fund portfolio to see how these new hedge fund products help to evaluate the rule-based strategy of investing into large hedge funds. Equation (9.1) gives us a general structure which simplifies to:[85]

$$\text{TOP50} = \text{alpha} + \text{beta*Index(DJCSI or HFRI)} \tag{9.2}$$

Table 9.5B tells us that over the period 2002–2010, alpha comes to 0.24% per relative to the DJCSI. But neither the DJCSI nor the HFRI is an investable proposition. Since the arrival of investable indices, we can replace them by their investable counterpart. For example, we can write for the DJCSI index:

$$\begin{aligned}\text{TOP50} = \text{alpha} + \text{beta*(tracking error}\\ + \text{(all-hedge-beta)*DJCS AllHedge Index)}\end{aligned} \tag{9.3}$$

$$\begin{aligned}\text{DJCSI} = \text{tracking error} + \text{(all-hedge-beta)}\\ \text{*DJCS ALLHedge Index}\end{aligned} \tag{9.4}$$

and

$$\begin{aligned}\text{Total Peer Group Alpha for TOP50} = \text{alpha}\\ + \text{beta*tracking error}\end{aligned} \tag{9.5}$$

Equations (9.2), (9.3), (9.4), and (9.5) give us a framework for evaluating the value of investing in a multi-manager hedge fund portfolio such as TOP50 (or a FOHF) compared to a passive, investable hedge fund index portfolio. For instance, the total peer group alpha in Equation (9.5) can be interpreted as the sum of a manager selection premium relative to the industry average (peer-group alpha) and a liquidity premium in that by investing into the Top 50 hedge fund managers investors are

[84] Unrealistic rebalancing rules amount to adding an additional liquidity premium to the index's return.

[85] Both the TOP50 and TOP50_2010 will recommend the same choice of funds going into 2011. The difference between the two lies with the expected alpha and beta relative to either DJCSI or HFRI.

accepting inferior liquidity terms compared to the DJCS All Hedge product. The existence of passive investable indices allows investors to determine how much of a multi-manager portfolio strategy's added value comes from superior manager choice and how much is just a liquidity premium.[86]

Finally, by comparing an investable index, which is a passive rule-based multi-manager portfolio, to an appropriate portfolio of LABs, investors can assess whether on average hedge fund managers add value to a portfolio of beta-bets. The hedge fund industry is maturing. The availability of performance data has substantially improved the quality and depth of performance benchmarks, leading to the development of passively investable index-like products. This encouraged academic research on hedge fund risk factors to develop which in turn sparked the development of beta-like hedge fund products adding to investors' risk management tools. In the next section we present some concluding remarks and conjectures on future development and areas of research interest, drawing from our observations on capital formation and product development trends.

CONCLUDING REMARKS

In the beginning of this chapter we posited an empirical metric for measuring the success of an asset management company—its ability to grow and retain investors' capital. Rooted in the world of private wealth management, confidentiality and privacy remain a common trait among hedge fund managers, many of whom do not release their performance data to the general public. Table 9.1 confirms that commercial hedge fund databases capture data from less than half of the hedge fund assets serviced by fund administrators and consistently so from 2003 to 2010. This raises an important and challenging question for researchers: *do the unobserved assets behave in a similar way to their observable counterparts?* We noted that the high degree of correlation between FOHFs performance with database dependent hedge fund indices hints at potential similarity between these two sets of hedge fund data—reporting versus nonreporting.[87] Nonetheless, until a more comprehensive answer to this question presents itself, the existence of such a large data gap casts a long shadow over the existing body of empirical hedge fund research. In this chapter we pull together empirical evidence and research from different sources to provide some insight to this important question. Specifically we infer from observable data a persistent concentration of industry AUM in the hands of a small number of large funds. In other words, success begets more success. It is a pattern that may well have been in existence from the early 1990s as shown in early studies by Fung and Hsieh (2000b) and Fung, Hsieh, and Tsatsaronis (2000). Table 9.4B reports a continuing divergence between the AUM of the top funds relative to the industry's median. Towards the end of 2010, the largest fund in the observable data universe is almost 1,000 times bigger than the median fund. The empirical evidence supports the conjecture that the unobserved AUM in the hedge fund industry is mostly in the hands of large hedge funds. It is for this reason we chose to develop our overview of hedge fund performance focusing on large hedge funds.

Over the past two decades, the hedge fund industry's path to success was punctuated by a number of landmark events, and followed the industry's evolution along this event-oriented time line. The convergence of risky bets in 1994 and the collapse of LTCM in 1998 were valuable lessons in risk management and portfolio construction. These events were touched upon and illustrate the systemic effect of funding risk has on leveraged strategies around the 2008 financial crisis. The arrival of commercial databases around 1994, incomplete as they may be, spurred much of the development in hedge fund benchmarks and academic research on hedge fund betas. One section is devoted to these topics—hedge fund strategy benchmarks and their inherent risk factors. Since then, investable hedge fund indices and funds that replicate hedge fund betas evolved.

Perhaps the most important event that shaped the modern day hedge fund industry was the arrival of institutional investors. This major shift in investor clientele had a profound impact on the supply of hedge fund products. The corporate governance requirement of institutional investors worked in favor of large hedge fund management companies shaping the capital formation trend of the industry since the turn of the century. We believe that the concentration trend of AUM in the hands of a small number of mega hedge fund management companies is pervasive. If this is the case, then the data gap we noted between AUM serviced by hedge fund administrators and AUM reported to commercial databases comes mostly from successful, large hedge fund managers who elected not to disclose data to databases. Insight on the properties of this data gap is not only important to models of hedge fund strategies and performance, but may potentially affect the future of hedge fund regulation. After all, regulating a small number of large management companies is quite a different proposition from regulating thousands of small private asset managers. However, if AUM concentration does continue without bounds, the probability of a convergence of risky bets like 1994 will rise, as increasing

[86] Investable indices typically carry lower fees than regular hedge funds.

[87] We note that FOHFs are not limited to investing only in hedge funds that are available in commercial databases.

amounts of leverage-able capital are concentrated in the hands of a few hedge fund managers. A convergence of opinions on certain risky assets could potentially lead to market disrupting events. More work is needed in this area of research.

We return to the discussion of analyzing performance. The continuing growth of industry AUM attests to the relative attractiveness of hedge fund strategies on average compared to their attendant risk (betas). The continuing success (in attracting AUM) of large hedge funds attests to the positive perception of value that large funds bring relative to their peers. The impressive AUM growth of the hedge fund industry has to be coincidental to delivering attractive performance relative to competing risky assets elsewhere in the global markets. To this end, Table 9.2A tells us that on average (based on reported indices like DJCSI and HFRI) the hedge fund industry returned statistically significant alphas adjusting for a broad range of risk factors. For investors searching for an alternative source of return to equities, Figure 9.4 provides justification for being attracted to hedge fund products over the period of 1996 to 2010. However, Table 9.2D tells us that alpha may be evaporating during the more recent period of 2002 to 2010. This is where large funds came to our rescue. We show that not only do the 27 large hedge funds used in the (Fung and Hsieh, 2000b) study deliver sizeable alpha relative to the eight risk factors (Table 9.3A), it also has substantial peer group alpha relative to the hedge fund indices (Table 9.3B). Therefore while alpha may have dissipated for the average hedge fund (or index) the same observation may not hold for the large hedge funds. We track the performance of the Top 50 hedge funds ranked by year end AUM. Although the passive strategy of investing in the past year's Top 50 funds consistently delivered alpha with respect to peer-group hedge fund indices, it failed to generate alpha with respect to the eight-factor model. However, the foresight-assisted portfolio of investing in the Top 50 hedge funds based on 2010 AUM rankings produced highly significant factor-based alpha as well as peer-group alpha. While this latter portfolio TOP50_2010 is not achievable without foresight, its performance illustrates that winners exist among large hedge funds, which is consistent with the belief that manager selection matters.[88] The fact that some large hedge funds can continue to outperform their peers reinforces the importance of better understanding the data gap from non-reporting hedge funds.

Finally, much has been written about the asymmetry in risk sharing between principal and agent inherent in variable compensation schemes of which hedge fund style incentive fees is one such variation. The problem arises when the incentive fee which hedge fund managers are entitled to, typically at 15–20% of new profits (or profits above a high water mark—HWM), ends up enticing a fund manager to take unreasonable bets. This can occur when the current loss-carry-forward (HWM) gets to a point that new profits only become viable when very large bets pay off. The cost of taking unreasonable bets is to further increase future loss-carried-forward should large bets fail. Faced with such a dilemma, a manager may be tempted to take an unreasonable risk so long as the option of closing shop and restarting in a new business (thereby resetting the loss-carryforward burden) if a large bet goes wrong is a feasible proposition. Historically, there are many incidents of fund managers re-emerging from a failed fund to raise fresh capital for new hedge fund ventures. However, closing shop and starting again clearly depends on the opportunity cost to the manager of closing a fund. It would be reasonable to assume that the larger the hedge fund management company, and the size as well as the age of the fund in question, the more costly it is for a fund manager to exercise the closure option—both in terms of the value of the track record of an operating fund and the reputation risk to the fund manager (or management company). The cost of fund closure being higher for large funds does partially mitigate this principal (investor) and agent (manager) conflict. In addition, from the early days of the hedge fund industry, investors have a preference for funds in which the agents (fund managers) invest a sizeable amount of their own wealth. This remains a common feature among successful hedge funds, one that helps align the interests of hedge fund investors and managers. This type of contractual arrangement has survived from the early days of the hedge fund industry. Taken together, these considerations tend to favor large hedge fund management companies and can provide some relief to the complex principal-agent conflict inherent in hedge fund compensation models. However, a formal model encompassing all of these features on how hedge fund managers are compensated has yet to emerge.

[88] It is also consistent with rational choices by investors for after all, it is investors placing more capital at risk with large funds that make them large funds in the first place.

References

Ackermann, C., McEnally, R., & Ravencraft, D. (1999). The performance of hedge funds: risk, return and incentives. *Journal of Finance*, *54*(3), 833–874.

Agarwal, V., Fung, W., Loon, Y. C., & Naik, N. (2010). Risk and return in convertible arbitrage: Evidence from the convertible bond market. *Journal of Empirical Finance*, *18*(2), 175–194.

Agarwal, V., & Naik, N. (2004). Risks and portfolio decisions involving hedge funds. *Review of Financial Studies*, *17*(1), 63–98.

Aggarwal, R., & Jorion, P. (2010). The performance of emerging hedge funds and managers. *Journal of Financial Economics*, *96*(2), 238–256.

Berk, J. B., & Green, R. (2004). Mutual fund flows, and performance in rational markets. *Journal of Political Economy*, *112*(6), 1269–1295 .

Bollen, N. P., & Pool, V. K. (2009). Do hedge fund managers misreport returns? Evidence from the pooled distribution. *Journal of Finance*, *64*(5), 2257–2288.

Bollen, N. P., & Whaley, R. (2009). Hedge fund risk dynamics: Implications for performance appraisal. *Journal of Finance*, *64*(5), 907–1037 .

Brave, A. P., Jiang, W., Partnoy, F., & Thomas, R. S. (2008). Hedge fund activism corporate governance, and firm performance. *Journal of Finance*, *63*(4), 1729–1775.

Brown, S. J., & Goetzmann, W. N. (2003). Hedge funds with style. *Journal of Portfolio Management*, *29*(2), 101–112.

Brown, S. J., Goetzmann, W. N., & Ibbotson, R. G. (1999). Offshore hedge funds: Survival and performance 1989–1995. *Journal of Business*, *72*(1), 91–117.

Brunmermeir, M. K., & Nagel, S. (2004). Hedge funds and the technology bubble. *Journal of Finance*, *59*(5), 2013–2040.

Cassar, G., & Gerakos, J. J. (2011). Hedge funds: Pricing controls and the smoothing of self-reported returns. *Review of Financial Studies*, *24*(5), 1689–1734 .

Duarte, J., Longstaff, F. A., & Yu, F. (2007). Risk and return in fixed income arbitrage: Nickels in front of a steamroller?. *Review of Financial Studies*, *20*(3), 769–811.

Edelman, D., Fung, W., Hsieh, D. A., & Naik, N. (2012). Funds of hedge funds: performance, risk and capital formation 2005 to 2010. *Financial markets and portfolio management*, *26*(1), 87–108.

Eichengreen, B., Mathieson, D., Chadha, B., Jansen, A., Kodres, L., & Sharma, S. (1998). *Hedge fund and financial market dynamics*. Washington, D.C. : International Monetary Fund.

Fama, E., & French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.

Fung, W., & Hsieh, D. A. (1997a). Empirical characteristics of dynamic trading strategies: The case of hedge funds. *Review of Financial Studies*, *10*(2), 275–302.

Fung, W., & Hsieh, D. A. (1997b). Survivorship bias and investment style in the returns of CTAs. *Journal of Portfolio Management*, *24*(1), 30–41.

Fung, W., & Hsieh, D. A. (2000a). Performance characteristics of hedge funds and CTA funds: Natural versus spurious biases. *Journal of Financial and Quantitative Analysis*, *35*(3), 291–307.

Fung, W., & Hsieh, D. A. (2000b). Measuring the market impact of hedge funds. *Journal of Empirical Finance*, *7*(1), 1–36.

Fung, W., & Hsieh, D. A. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies*, *14*(2), 313–341.

Fung, W., & Hsieh, D. A. (2002). The risk in fixed-income hedge fund styles. *Journal of Fixed Income*, *12*(2), 1–21.

Fung, W., & Hsieh, D. A. (2004a). Extracting portable alphas from equity long-short hedge funds. *Journal of Investment Management*, *2*(4), 57–75.

Fung, W., & Hsieh, D. A. (2004b). Hedge fund benchmarks: A risk based approach. *Financial Analysts Journal*, *60*(5), 65–80.

Fung, W., & Hsieh, D. A. (2006). Hedge funds: An industry in its adolescence. *Federal Reserve Bank of Atlanta Economic Review*, *91*(4), 1–33.

Fung, W., & Hsieh, D. A. (2009). Measurement biases in hedge fund performance data: An update. *Financial Analysts Journal*, *65*(3).

Fung, W., & Hsieh, D. A. (2011). The risk in hedge fund strategies: Theory and evidence from long short equity hedge funds. *Journal of Empirical Finance*, *18*(4), 547–569.

Fung, W., Hsieh, D. A., Naik, N., & Ramadorai, T. (2008). Hedge fund: Performance, risk and capital formation. *Journal of Finance*, *63*(4), 1777–1803.

Fung, W., Hsieh, D. A., & Tsatsaronis, K. (2000). Do hedge funds disrupt emerging markets? In: Litan, R. E., & Santomero, A. M. (Eds.), Brookings-Wharton Papers on Financial Services, pp. 377–401.

Getmansky, M., Lo, A. W., & Makarov, I. (2004). An econometric model of serial correlations and illiquidity in hedge fund returns. *Journal of Financial Economics*, *74*(3), 529–609.

Goldman, M. B., Sosin, H., & Gatto, M. A. (1979). Path dependent options: Buy at the low, sell at the high. *Journal of Finance*, *34*(5), 1111–1127.

Hajim, C., & Lashinsky, A. (2007). How bear stearns lost its way, fortune, August 21st, 2007 at http://money.cnn.com/2007/08/20/magazines/fortune/bear_stearns.fortune/index.htm.

Khandani, A., & Lo, A. (2007). What happened to the quants in august 2007?. *Journal of Investment Management*, *5*(4), 5–54.

Khandani, A., & Lo, A. (2011). What happened to the quants in August 2007?: Evidence from factors and transactions data. *Journal of Financial Markets*, *14*(1), 1–46.

Kosowski, R., Naik, N., & Teo, M. (2007). Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics*, *84*(1), 229–264.

Lowenstein, R. (2001). *When genius failed: The rise and fall of long-term capital management*. New York : Random House.

Merton, R. (1981). On market timing and investment performance I: An equilibrium theory of value for market forecasts. *Journal of Business*, *54*(3), 363–407.

Mitchell, M., & Pulvino, T. (2001). Characteristics of risk in risk arbitrage. *Journal of Finance*, *56*(6), 2135–2175.

NACUBO (National Association of College and University Business Officers), Endowment Study, 1998–2008.

NACUBO (National Association of College and University Business Officers)—Commonfund, Study of Endowment, 2009–2010.

Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Palazzolo, C. (2009). *The convertible bond market dislocation of 2008: Creating opportunities in 2009 and beyond*. White Paper: AQR Capital Management LLC.

Patton, A. (2009). Are 'market neutral' hedge funds really market neutral?. *Review of Financial Studies*, *22*(7), 2496–2530.

Patton, A., & Ramadorai, T. (2012). On the high frequency dynamics of hedge fund risk exposures. *Journal of Finance* (forthcoming).

Pensions and Investments. The P&I 1000. 1998–2010.

Sharpe, W. F. (1992). Asset allocation: Management style and performance measurement. *Journal of Portfolio Management*, *18*(2), 7–19.

Stulz, R. M. (2007). Hedge funds: Past, present and future. *Journal of Economics Perspective*, *21*(2), 175–194.

10

Performing Due Diligence on Specific Managers and Funds

Learning Objectives

After completing this reading you should be able to:

- Identify reasons for the failures of hedge funds in the past.
- Explain elements of the due diligence process used to assess investment managers.
- Identify themes and questions investors can consider when evaluating a hedge fund manager.
- Describe criteria that can be evaluated in assessing a hedge fund's risk management process.
- Explain how due diligence can be performed on a hedge fund's operational environment.
- Explain how a hedge fund's business model risk and its fraud risk can be assessed.
- Describe elements that can be included as part of a due diligence questionnaire.

Excerpt is Chapter 12 of Hedge Fund Investing: A Practical Approach to Understanding Investor Motivation, Manager Profits, and Fund Performance, *Second Edition, by Kevin R. Mirabile.*

This chapter is concerned with the process of performing due diligence on a specific manager that is being considered for investment. **Due diligence** is the term used to describe the process of evaluation and analysis that an investor follows to get comfortable with a strategy, a manager, and a fund prior to making an investment. It includes all the steps that are needed to get to know why and how a fund came into being; the skills its founders or current partners claim to have mastered; the evaluation of the timeliness, accuracy, and consistency of manager and fund information; the reliability and independence of service providers; and far more. The chapter is designed to describe some of the common procedures used to investigate a manager's trading and investment skill; ability to manage operations, credit, and liquidity risk; and importantly, the ability to run a business.

Investors would be wise to assess all three skill sets needed to run a modern-day hedge fund—investment, operational, and business skills—before committing capital to any fund. Those who do lower the risk of surprise and improve the prospects of getting what they anticipated from any opportunity. Managers today are rarely good at all three things. In fact, due to the changing demands of institutional investors and the failure of 2008, it is most often impossible for any one person or leader to process a high level of competency in all three areas. This has profound implications for how hedge funds are run today.

Founders who are the CIO and CEO and who also manage a fund's risk and business model themselves are, quite frankly, considered dinosaurs today. Today's top managers are very often making the choice to either be the CIO and share the CEO role or, where they are serving both roles, to bring in senior partners who can add depth and have specialized risk, accounting, operations, or trading skills beyond their own. Investors in hedge funds now are quite focused on making sure the organization can provide all three skills at a very high level, even though any one founder cannot. Many firms with great investment ideas lack risk controls or business skills. Others who are control oriented and can run a business may not be able to generate the returns that are needed. Finding the best combination of return and infrastructure without excessive business risk is sometimes more of an art than a science.

Investors need to think about performing due diligence in a way that is comprehensive and holistic. This means that it must cover the investment process, operational environment, and business acumen of a manager and a fund. It should also be both quantitative and qualitative. Quantitative analysis ensures that every aspect of due diligence is covered and nothing obvious gets missed. This is sometimes referred to as the checklist approach. Qualitative analysis is much more investigative and allows investors to ask the type of open-ended questions needed to assess the attitudes and culture of a manager and the consistency of message or a process throughout an entire organization.

This chapter organizes the due diligence process into four sections:

1. Investment process
2. Risk management
3. Operational environment
4. Business model assessment

In each section, we cover a few of the more important quantitative or check-the-box criteria, as well as those qualitative and subjective criteria that very often help an investor expose a manager's underlying strength or weakness.

However, before diving into the investment, risk management, or operational or business model risk assessment for any manager or fund, a few words of caution are worth noting. Each of the following axioms provides a word or two of caution about the due diligence process. If they are followed, investors will at least position themselves for the best possible outcome, given any set of idiosyncratic circumstances or market outcomes that may arise. Hopefully, they will eliminate at least some of the less fortunate surprises that inevitably occur when investing in lightly regulated private vehicles such as hedge funds.

10.1 BE PREPARED

First and foremost, an investor must be well prepared when trying to assess any investment opportunity. This is particularly true in trying to evaluate the skill of a hedge fund manager or the attractiveness of any investment process. Being prepared means having a firm understanding of why you are considering a particular type of investment style or strategy. It means knowing how that style should react to various market conditions. By fully understanding the strategy beta, you will be able to focus more time and attention on the manager's ability to generate alpha relative to peers in the same strategy.

10.2 LEARN FROM THE PAST—FROM BOTH SUCCESSES AND FAILURES

Getting caught up in the past success of hedge fund investing is quite easy. In fact, very often an investor's previous success or past association with an industry legend can create just enough hubris to get the investor in trouble in the future.

00011693-00000002_CH10_p187-200.indd 188 05/07/22 3:08 PM

Investors like to believe they have found the next George Soros, Jim Simons, or Paul Tudor Jones. They rarely like to mention when they end up having found the next Madoff or allocated capital to yet another fraudster with supposed investment superiority or acumen.

So why do funds fail? There are many reasons. Investors should remind themselves of those reasons and seek answers about previous fund failures in the strategy they are considering as part of the due diligence process. Funds can fail as a result of bad investment decisions. Funds can make several compounded bad decisions or just a few concentrated calls on the markets or individual securities that perform very poorly. Funds can also fail due to all sorts of frauds, including accounting frauds, valuation frauds, or misappropriation of funds. Funds can fail due to excessive leverage, improbable probabilities, unexpected events, and tail risk. Funds can fail due to a flood of unanticipated withdrawals of capital at the least opportune time. Funds can get caught in squeezes by the street or by other hedge funds. Funds can fail as a result of a lack of supervision or compliance controls related to insider trading. Funds can fail because of their own actions or the acts of others. A prime broker like Lehman or MF Global can go bankrupt and take a fund with it. Funds can fail when liquidity dries up, and they can't meet redemptions and must sell into a market that no longer exists, at least at that moment.

Investors need to evaluate each strategy and each manager with open eyes and take great care not to be blinded by the past success of any one person or strategy. The more an investor assumes things can and will go wrong, the better the due diligence process and the more likely it is that an investor will uncover problems and/or ensure that attractive opportunities have a higher probability of success.

10.3 IF IT LOOKS TOO GOOD TO BE TRUE, IT PROBABLY IS

The individual employees who are performing investment due diligence should always remember to trust their own judgment. If it looks too good to be true, it probably is. Too often, firms make allocation decisions by committee. In this setting, it is easy for one voice, perhaps one's boss, to direct the decisions of the team. Individuals may feel too intimidated to speak out or to communicate details that may reflect negatively on a manager who is well known to the firm or who has established a high-level relationship with the CIO or other members of the organization. It is essential that people who come across facts, figures, opinions, references, or any other relevant information uncovered during their work escalate their concerns. If something in a manager's track record, background, or pedigree comes into question, analysts must raise their hands and speak their minds. Very often, the most notorious frauds and funds that were behaving badly were quite obvious to those in direct contact with the manager. Often only in retrospect is everyone able to see what should have been obvious to all. In the end, no matter how uncomfortable, escalation of questionable information will benefit the organization and will be appreciated. If things don't feel right or just don't pass the smell test, then they should be voiced within the research team before ever getting to the investment committee for allocation decisions or approval.

10.4 REMEMBER, IT'S STILL ABOUT RETURNS!

So much media and regulatory focus has been placed on the need for hedge funds to raise their standards and adhere to best practices, particularly related to operations, credit risk, and liquidity, that it is entirely possible for some investors to actually overemphasize some aspects of the due diligence process. Investors may place so much emphasis on operational due diligence alone that they end up picking managers with great controls that actually don't add much return to the portfolio at all. Investors should be mindful not to overweight infrastructure and business model risks. There needs to be a good balance between the investment opportunity and the risk profile of the fund. A set of minimum standards or bright lines is helpful to ensure that the balance doesn't shift too far in the other direction, resulting in the selection of managers with great strategies and poor controls or business risk. The need for independent administration, audits, and top-tier service providers is generally not negotiable. However, infrastructure and business models should be strategy and life cycle specific. A recently launched equity variable bias fund operating in the United States may not need multiple prime brokers, real-time disaster recovery, and a hot backup or succession plan, certainly not all on day one! The reason investors invest is to increase return and to reduce risk. As simple as that sounds, it can easily be forgotten.

10.5 COMMON ELEMENTS OF THE DUE DILIGENCE PROCESS

The due diligence process today is very different than it was in the past. In the past, manager reputation and performance were the most important factors. Investors did little digging into the how and why of performance and the safeguards in place to

protect assets. This was due in part to the lack of leverage that individual investors had with the managers. They were often considered "lucky" just to have gotten into a well-known fund with limited capacity. Institutions had relatively small exposures to hedge funds at the beginning and needed to have a high return to have the investment matter to the portfolio. Finally, managers were rather selective and if you asked too many questions you would simply be told to go elsewhere.

As the industry matured and as more institutions came into the market, not to mention some very high-profile frauds, the due diligence process expanded. Today, both managers and investors spend a great deal of time trying to learn where a manager's "edge" is coming from and that their investment is safeguarded and properly valued.

The due diligence process involves two separate but closely related evaluations. One evaluation is of the firm's investment process and related risk controls. The other is related to the fund's operations and business model. Today both are considered almost equally important and are often interrelated.

10.6 INVESTMENT MANAGEMENT

This section is a high-level summary of some of the common themes and questions investors use to perform due diligence on a fund's investment process. It is by no means comprehensive and is intended to provide a sample of the types of issues that investors are facing when trying to determine the value and quality of a fund's investment process themselves, independently from the documentation and high-level representations of senior management. Most if not all of the questions related to a fund's investment process should be done in person with as many people of mixed seniority as possible. The goal is to find out what is really going on at the fund and not just have a pitch book recitation by the investor relations staff. Remember that if you cannot get access to the actual decision makers when you are trying to invest, it is unlikely you will have access if something goes wrong! Hearing it directly from the firm's founder and lieutenants is always best, but getting the same message from a diverse number of people who perform certain tasks on a day-to-day basis or should be familiar with them is also quite useful.

What Is Your Strategy, and How Does It Work?

Investors who are evaluating a manager often start with high-level questions that provide them with some context about the firm, its investment strategy, and how it works.

1. What is the manager's self-described style, and how does it fit within a particular classification scheme?
2. What are the current themes included in the portfolio, and what are the fund's highest convictions or most concentrated positions?
3. How has the portfolio evolved over the past several quarters, and what is the outlook for any changes, given current market conditions?
4. Does the firm manage the fund to specific gross and net exposure targets, and if so, where does the fund stand today versus those targets and why?
5. What is the portfolio turnover and number of days to liquidate, and what are the triggers that result in a reversal and a sell or buy to cover or exit decision? Are the triggers hard coded into the risk management process or just guidelines?
6. How are stop losses used to manage risk? Are they used at the position level, portfolio level, or both? Are they always executed or are there exceptions?
7. How quantitative is the investment process, and how much does the firm rely upon proprietary models? If so, what are they based on, and how are they developed, back-tested, and allocated capital, or decommissioned when failing to perform?
8. How are short sales used—as hedges, alpha generators, or both? Has the firm ever been hurt by a short squeeze, recall, or buy-in on borrowed shares?
9. How are listed or OTC derivatives used in the portfolio?
10. How is trading organized? Does the firm have a central trading desk for all external order flow?
11. Is the strategy capacity constrained?
12. Has the fund ever held private investments, and if so, why and how do they support the core investment strategy?
13. Is the firm more return oriented, asset growth oriented, or both?
14. How does it manage the conflict between generating returns for existing investors versus growing the assets under management?
15. Has the fund ever been soft or hard closed or returned capital to investors?

Once this initial set of questions about the firm and the fund has been answered, the investor can begin to drill down into more specific questions about the firm and the process followed to manage investing.

How Is Equity Ownership Allocated among the Portfolio Management, Trading, and Research Teams?

Understanding the firm's ownership structure and how things get done is critical. The participation of the investment team in some form of ownership is a critical area of differentiation among firms. Some firms do not share equity ownership among the portfolio managers or traders in the firm. Other firms use equity ownership as a key feature of their professional talent retention program and to attract and groom new talent for future leadership positions. One model is not necessarily better than the other. Each has its pros and cons. Investors often have a view one way or another. The key thing is to understand the firm's philosophy and how it impacts performance, talent acquisition, and retention.

Is the Track Record Reliable?

After an initial set of questions about the firm, an investor will want to dig a bit deeper into the manager's and the specific fund's track record. According to a recent report on hedge fund due diligence performed by the not-for-profit organization the Greenwich Roundtable, investors should inquire whether the track record is comparable to similar strategies, has been audited, and is long enough for statistical evaluation and inference, whether returns were impacted by fund size, and if the team that produced the historical track record is still in place today. Additional questions about the track record should include how it performed during periods of market stress and how it relates to the portfolio manager's experience at previous firms, if applicable.

Who Are the Principals, and Are They Trustworthy?

Investors who are thinking about a particular manager need to allocate resources to both references and background checks. Newer managers are a bit trickier than established ones. New managers need to have references from previous hedge funds or banks verified, both those that are on the manager's reference list and those that the investor can obtain independently. More established managers won't need to be interrogated about previous firms or employers from the distant past as much as start-ups or firms lacking a reliable track record. When interviewing established managers or talking to their references, focus the dialogue more on the manager's motivation, behavior during crisis or stress periods, ability to communicate to investors, and experiences other investors have had, including both positive and negative over the life of the fund.

The ability to verify a new manager's prior employment is just the beginning. Did they do what they described? Were they the trigger puller or acting in a support role? Did they generate ideas or implement ideas? Did they operate individually or within a group? Often, the best way to find out is to ask former colleagues, partners, managers, clients, or other independent parties. Investors performing reference checks can think of the process as a 360-degree review of what a person has claimed about the past or even about his or her current skill set. If enough people are saying the same thing about a person or a firm, then an investor can take some comfort. If significant differences of opinion about a person or a manager are uncovered, they should be thoroughly vetted and then discussed directly with the manager. Most often, there is a good explanation—but not always!

Hedge fund scandals and blowouts over the past 10 years have also changed the way investors need to think about the people they are entrusting with their assets. The high-profile blowups of several hedge funds, including Bayou Management and Madoff, have raised important questions about what can be done to uncover fraud in advance of getting hurt.

Background checks can certainly be obtained. Most investors can get a pretty comprehensive report on a manager for as little as $750. More depth requires more money, but in some cases, it certainly may be worth it. Very often there is information just below the surface that can be used to discover managers who have had trouble in the past.

In many hedge fund frauds, investors could simply have searched public databases and found litigation, criminal accusations, or civil disputes that may have at least been used to raise questions about the manager's integrity or past deeds. Many frauds simply go undetected because of a lack of proper due diligence. After the fact, it is easy to point to smoking guns or indicators that something may have been wrong.

Either on their own or using independent resources, investors should always consider calling former employers, track down credit reports, and contact auditors, administrators, and prime brokers to verify what a manager is representing in fund documents or through verbal communication. An investor should include searches of websites maintained by the SEC and read the manager's Form ADV in detail before deciding to invest. Information about bankruptcy, federal or appeals court records, or other sources to uncover information about a manager are readily available.

Investors should always look for related party activity when doing background searches. Does the manager or the principal own other business activities that do business with the fund? A fund that uses affiliated brokers or administrators can lead to real problems, especially if the manager did not disclose it in advance.

Investors need to make sure they have access to the people at the top of the firm. It is also important that investors deal directly with the actual risk takers and decision makers and not just investor relations, salespeople, or junior staff at the fund.

10.7 RISK MANAGEMENT PROCESS

Evaluating the risk management process and procedures entails hybrid questions that should be answered by all levels of management of the firm. Portfolio managers and traders, as well as operations staff and risk managers, all have valuable insight for investors. Risk management–related questions should be asked as part of both the evaluation of the investment process and the operational environment of the firm.

How Is Risk Measured and Managed?

Risk management is an emerging discipline at many funds. That is not to say risk was not managed well in the past. It is more a reflection of the fact that risk measurement and reporting and the decision to take action is evolving toward a more independent model that segregates risk taking and investing from risk measurement and management. Today, many funds have dedicated risk managers who report to the CIO or the CEO independently from the portfolio managers and traders. Many firms also employ independent risk service providers to report risk to investors completely independently from the firm. Some use fund accountants and administrators to achieve this goal.

Investors must first inquire about the potential risks that a strategy will expose them to in the normal course of business and then inquire about the additional idiosyncratic risks presented by a specific manager within a strategy.

Investors next might want to ask if there are written policies and procedures to monitor and measure each of the applicable risks and whether there is a risk committee to which each of the measured risks gets reported on a daily, weekly, or monthly basis. Is there a risk management culture in the firm, and do traders and portfolio managers, as well as back-office and support staff, understand and participate in risk management and mitigation?

An evaluation of a fund's risk and its process to measure and monitor risk inevitably morphs into a discussion of systems and technology. Modern-day hedge funds use complex processes to originate and control risk. Investors should inquire about the process used by the firm to choose a single or multiple sets of risk platforms and the consistency of risk measurement between the traders, portfolio managers, and investors.

Great care should be taken to understand the inputs and assumptions used in the firm's risk models and whether the portfolio is stress-tested on a regular basis to measure the impact of changing assumptions. Too often, traders use a different system with different inputs and outputs to monitor their risk than the firm is using to report risk to investors. This can result in a significant miscommunication and lead to problems when things go wrong.

Importantly, each style of hedge fund strategy has a unique and sometimes mutually exclusive set of risks. An equity fund has beta exposure, whereas a credit arbitrage fund may have credit spread exposure. Event funds may be exposed to very specific catalysts that can have a significant impact on performance. Global macro funds have more inflation, interest rate, and currency exposure than most single-strategy funds that trade a specific asset class. The information made available to investors needs to cover both the generic risks common to many funds and the particular risk associated with a single strategy and a specific fund.

How Are Securities Valued?

A firm's valuation policy is another critical area that investors need to examine when considering a fund. What percentage of the fund's assets are exchange traded and marked to market via exchange prices versus model prices or broker quotes? Does the administrator take responsibility for valuation, or does the manager retain responsibility? Who can override prices, and is there a formal documented process to do so?

What Is the Portfolio Leverage and Liquidity?

An investor needs to evaluate the current and historical changes in leverage, the sources of leverage, and the liquidity of the fund's portfolio over time. In doing so, understand where the particular manager may deviate from peers or exhibit leverage or illiquidity that can cause performance to deviate from the expectation of the strategy. Some strategies such as global macro have fairly uniform leverage terms and liquidity based on the products they trade. Others, such as fixed income and convertibles, use varying degrees of leverage and have very different liquidity profiles from firm to firm. An investor who expects

to earn the mean return of an index, such as a convertible index, can get very different results depending on the manager's leverage and liquidity or orientation. Depending on the answers to questions about leverage and liquidity, investors may need to adjust their expectations for returns that were initially based on the strategy or a comparable index. Investors should inquire how the current leverage and liquidity in the existing portfolio compares to previous quarters. Liquidity will also have a direct effect on a fund's capacity and ability to handle larger amounts of investor capital.

Does the Strategy Expose the Investor to Tail Risk?

Certain strategies may expose investors to unanticipated tail risk. Investors need to perform their own analysis of a fund's data and determine its skewness or kurtosis. They should also inquire whether the manager believes that tail risk exists and, if so, get an explanation of how it is hedged or whether investors need to accept or hedge it on their own.

How Often Do Investors Get Risk Reports, and What Do They Include?

Investors are always entitled to periodic reporting from the fund. The fund offering documents and materials normally state explicitly when and what is provided. Some funds report on risk using a standard package produced by a third-party risk provider or include key risk statistics in their monthly fact sheet or periodic investor letter. Investors should obtain this information, ensure they completely understand it, and compare it to peers and/or strategy-level indices for consistency and to identify unique risk taking by the fund prior to investing.

Do the Fund Terms Make Sense for the Strategy?

Investors want to assess if the terms make sense for the strategy that is being offered. A long-only manager getting a 2 and 20 fee or a liquid long and short equity strategy with a one-year lock-up can be red flags. Investors can compare any specific manager or fund to its peers to see if the terms make sense. Law firms, accountants, and many commercial databases are good sources to collect comparative data on fund terms. Investors who give away terms such as lock-up to managers are not getting compensated for the risk they are taking. Investors who pay high fees for market beta are overpaying for something that could be more cheaply replicated on its own.

Investors should inquire about the fees, high-water mark, and hurdle rate related to any investment. Is the fee appropriate and in line with peers? How is the hurdle rate calculated and by whom? Is the high-water mark reset annually, or is it perpetual? Does the portfolio's liquidity match the liquidity offered to investors, and if not, is the gap a reasonable one? Is there a lock-up period before redemptions are allowed? Can the fund gate or suspend redemptions?

10.8 FUND OPERATING ENVIRONMENT, DOCUMENTATION, FINANCIALS, AND SERVICE PROVIDERS

This section provides a high-level summary of some of the common themes and questions investors use to perform due diligence on a fund's operational environment. This includes validation of a fund's internal procedures and its relationships or exposure to important service providers. Once again, it is by no means comprehensive and is intended to provide a sample of the types of issues that investors are facing when trying to independently assess the quality of a fund's operating environment.

An operational due diligence program includes review of several important aspects of how a hedge fund is managed and how its interaction with the fund itself is controlled. The primary purpose of operational due diligence is to ensure that no significant additional risk of loss is being created for investors related to the settlement of securities, process of corporate actions, misappropriation or theft by employees or agents, or any other breakdown in the manager's confirmation, verification, valuation, and reconciliation process.

An investor performing operational due diligence generally focuses on assessing the adequacy of a manager's internal controls, consistency of fund documents and legal representations, and the risks of loss due to counterparty or service provider failure.

Internal Control Assessment

A review of a manager's control environment includes many items. Some of the more important items that investors can review include the qualifications of people at the fund, the quality of the written procedures, and the ability of the team to execute them each day and clear any breaks of exceptions that may occur, as well as the fund's exposure to derivative counterparties and the protections provided by its governance structure. It is not enough to have a good plan. The people must be qualified, and the process must be followed every day.

Let's start with a brief discussion of the qualifications of the people who run the firm and are empowered to act on behalf of the fund. Does the CEO support a culture of control and compliance, or is he a person who doesn't really like to follow the rules? The message from the top down needs to be one of safeguarding assets, where following a process is supported and, more than anything, investors are valued and treated with respect. An assessment must be made whether the operations, accounting, treasury, technology, compliance, and other personnel are truly qualified for the positions they hold and the products the fund is trading. Investors must ask whether the managers have experience managing and whether they themselves are qualified to get the job done correctly. It is not unusual for background checks to be done on the firm's COO, CFO, accounting managers, or other key back and middle office personnel, in addition to the checks normally done on the investment and research team members. Does the firm hire people with experience, and how do they train their staff and maintain and grow their skill set and knowledge base outside the big-firm environment from which most people came?

A review of the fund's written procedures related to trading, derivatives, cash and securities processing, and position servicing can be performed during a site visit to the firm. It is now routine for investors to inspect documents outlining firm procedures and evaluate those procedures to see evidence of how the written policies are being followed. Unfortunately, these sorts of documents and process reviews are not really comprehensive or efficient. In reality, it only gives investors a sense of whether the management of the hedge fund takes the whole control process seriously. Some hedge fund managers have made real progress in this area and have gone as far as to get an outside audit firm to evaluate their procedures and controls on a periodic basis and issue an opinion of whether the controls are sufficiently well designed and have been tested and are operating effectively.

Compliance is another critical area of investigation. Most firms today have either their own in-house compliance function or an outsourced relationship with a compliance service provider. All but the smallest of firms have some resources dedicated to compliance. Compliance practices that can be verified include the existence of a code of ethics, prohibition of related-party transactions, and restrictions related to employee trading.

Counterparty risk related to the use of OTC derivatives and other trade counterparties is now a topic routinely covered in the operational due diligence function. Funds have varying exposure to the firms they trade with. Firms holding up-front margin have failed and closed out fund positions without any return of the margin payment held in the account. The Lehman, MF Global, and Refco failures all resulted in hedge funds losing money that belonged to the funds they managed. When a fund has cash being held with a dealer and the dealer goes bust, it may not get the money back—ever! At a minimum, there may be a lengthy court battle, and then recovery may only be 30 or 40 percent of the amount they thought was safe and secure as cash on the balance sheet or a receivable from a dealer. Hedge funds have to diversify this risk among several firms, take action to move any excess cash out of firms that may be at risk, and have a process in place to monitor the risks of these counterparties every day.

Finally, is there a governance structure that extends beyond the CEO? If so, does it have any teeth? Increasingly, investors are holding fund directors and advisors accountable for the decisions a fund manager makes related to service providers and the actions that a fund takes that lead to style drift or even fraud.

Documents and Disclosures

An investor needs to verify with the listed law firm in the fund documents that they were responsible for the original content and for any updates. Investors generally rely on the quality of the law firm that has created or amended fund documents. On some occasions, managers have been known to make changes to their fund documents without the consent or even knowledge of the law firm that drafted the documents. An investor can actually check the document to see if any changes were made after the as-of date on the face of the document. Changes made after that date should be discussed with the manager immediately and with the law firm. In addition, verify that the law firm cited in the fund documents is still under a retainer agreement with the manager or fund. Most law firms will at least indicate whether they currently represent the fund.

Next, it is very important to ensure that the fund offering memo, subscription agreement, limited partnership agreement, investment management agreement, Form ADV, and website are all saying the same thing at the same point in time. Very often, these documents can change after a fund launch, and some may no longer match the terms of the offering memo that ultimately governs an investor's rights and obligations. Particular attention must be paid to fees, liquidity, side pockets, gates, suspension rights, creation of subfunds or SPVs, and change in offshore investment managers.

Investors in hedge funds must carefully consider the conflicts of interest section of the fund's offering memo (OM). This is where the law firm drafting the documents most clearly states any concern about nonstandard arrangements with managers, co-investors, and others. Be sure to query these representations,

especially if they are vaguely worded, such as "certain principals or affiliates of the manager may maintain relationships with certain fund counterparties but will always operate on an arm's-length basis." This can mean anything from "our administrator also does some accounting for Citibank, where we hold our operating accounts" to "the manager takes a personal rebate from the prime broker."

Caution is advised when there are either insufficient or extremely broad risk disclosures. Insufficient risk factors are more of a red flag than too many. In addition, overly broad and irrelevant risk factors are also a red flag. In the latter case, the law firm may not have adequately looked at the manager's program or is drafting the OM so broadly that it is looking primarily to protect itself, not the investor.

Also be sure to check registration and compliance language with independent counsel specializing in the regulation of the strategy mix that the manager trades. Investment advisors may or may not cover commodities. Foreign advisors may not be exempt from U.S. registration, even if you are investing through a feeder fund. The scope of exemptions from registration is constantly changing and should be current.

Document review should include a complete read of all the fund documents to ensure that the terms are reflective of the discussions the investor has had with the manager. Very often, terms are omitted or not discussed in person, yet later are uncovered at the last minute when documents are being reviewed or, worse yet, being completed. Redemption right, liquidity, notice period for redemptions, hard or soft lock-ups, early redemption fees, the ability of the manager to gate or generally suspend redemptions, and the amount and timing of payouts following redemption (there may be a 5 to 10 percent holdback pending annual audit) should all be confirmed with the manager verbally and reconciled to the fund documents, The same holds true for management and incentive fees subject to a high-water mark or hurdle rate and the expenses that a fund will bear related to start-up costs, market data, or other costs. Finally, subscription rights (timing of subscriptions, minimum subscription amounts, limits or commitments on capacity) should be reviewed and agreed upon with the manager.

Other important considerations when reviewing fund documents include the powers of the manager: Are they very broad or relatively narrow? Are there any restrictions related to the use of leverage or concentration in the documents? Can the manager amend fund documents, such as the limited partnership agreement? What are the key man event rights or notice to investors? Does the fund have indemnification provisions, and to what extent does the fund itself indemnify the manager and any directors? Typical indemnifications should not extend to the gross negligence, bad faith, fraud, or willful misconduct of the manager.

The documents should also clearly state the manager's obligations for reporting to investors, including audited financial statements and any tax implications associated with the fund's investments that can impact fund investors.

Financial statements are also important documents for investors to evaluate when considering a particular fund. Analysis of the fund's last financial statements can provide investors with valuable information about the manager and the fund. A review of the financial statements should start with reading the opinion and confirming that it is, in fact, "unqualified." This basically means the independent auditor has reviewed the company's records and issued the statements without any material caveats or concerns.

An investor can examine the balance sheet and income statement to see if it makes sense based on the fund's trading strategy. Equity funds should look very different than global macro funds or fixed-income funds. An equity fund that purports to have a low level of leverage that has been declining year over year should not report a high level of interest expense that is rising year over year in its financial statements. A global macro fund most likely has interest income earned in its income statement. Investors can quickly learn to expect certain patterns in the balance sheet and income statements of funds pursuing specific strategies. If a particular fund doesn't make sense, then questions need to be asked immediately. Leverage on the balance sheet can be recalculated and compared to the leverage expected by the strategy or represented by the manager. In addition, funds that have a long-term buy-and-hold strategy should not be generating large amounts of realized gains or losses each year; they should have sizable unrealized gains or losses instead.

Any unusual line items that raise a red flag should be discussed with the manager or even the auditor. Material items that are unique or different are usually explained in the footnotes. Reading the footnotes is critical, as it is often where the really important items get clarified, such as the use of derivatives or litigation.

Investors can and should recalculate the fees paid by the fund to the manager and make sure they make sense. Fees should be compared to the pitch book and other fund documents such as the OM. There should never be incentive fees earned in years when the fund lost money.

Finally, check the equity section to see if the general partner is continuing to invest in the fund. Withdrawals or reductions in equity are a big red flag. Investors should thoroughly discuss

00011693-00000002_CH10_p187-200.indd 195 05/07/22 3:08 PM

and understand the changes in the capital accounts of the general partner and any significant key people who run the fund.

Service Provider Evaluation

Investors should expect that a hedge fund will make all of the key contacts at their service providers available to them so they can verify the scope of any services being provided. Investors can also obtain internal control letters and audited financial statements from the fund's service providers to make sure there is an independent check on the service providers themselves. It is not uncommon for investors today to interview service providers and discuss the role they play in executing trades, providing technology, performing valuations or verification, and safeguarding assets.

10.9 BUSINESS MODEL RISK

This section of the text deals with the risks encompassed in simply running the business of being a hedge fund. It is an entrepreneurial activity that is now being held to institutional best practices as standards. It is a business that has risks that are similar to many others. Will there be adequate cash on hand, where does the working capital come from, and is it organized properly? Is there a succession plan? What happens if too many investors redeem at the same time? These and many other questions are relatively new to hedge fund managers. During the early stages of the industry's growth, hedge fund managers rarely had to face these sorts of issues. They benefited from relatively low barriers to entry, and much of the time there were even relatively low barriers to success. Capital was plentiful, leverage was available, markets were rising, rates were low, and spreads were narrowing!

Today, there are much higher barriers to entry, and the barriers to success have gotten much higher. Not everyone succeeds. In fact, over the past several years, as many as or in some cases more funds have failed than have been launched. This rarely occurred in the past, if ever. The implication to investors can be significant. A fund that fails needs to be liquidated, often in adverse conditions. Capital may be tied up; worst case, there may be litigation and embarrassment or a loss of confidence in the investor's decisions. No investor wants to give money to a manager who then closes suddenly, having run out of cash to run the business. Many hedge fund managers are simply not prepared for this brave new world. They often have never needed to develop the business building, financial planning, and strategic planning skills needed to launch a business, retain talent, and grow a business over an extended period of time.

A white paper written in 2011 by Merlin Securities, a prime broker that caters to hedge funds managing less than $1 billion, captures many of the challenges that managers are facing today with respect to the business of running a hedge fund. The executive summary reads:

> 2010 was a transformative year for the hedge fund industry and served as a strong reminder that managing money is not the same as running a business. The significant number of small, mid-size, and large fund closures already in 2011 provides continuing evidence of the material, multifaceted challenges facing operators of hedge fund businesses. Managers who understand the distinction between managing money and running a business and who execute both effectively are best positioned to maintain a sustainable and prosperous business—to achieve not only investment alpha, but also enterprise alpha.

Managers who take business model risk seriously and who are actively modifying their business to adapt will be successful. Those who do not adapt will not be successful. Given the massive changes in this industry over the past five years, it is virtually impossible for a manager to simply stand still and still be successful.

Since 2008, understanding the cost base and the revenue and expense model of any fund has become critical to predicting whether a fund will survive. Managers need to spend a great deal of effort to design a sustainable business model that can survive under a wide range of performance scenarios and that can support the ever-increasing demands of investors. Not all funds are able to do this well. In fact, many funds have decided to simply return capital or close rather than evolve to meet the demands of the market and of their changing investor base.

A manager's ability to control costs and predict revenues, while not easy, is essential for a firm to survive. Firms that overrely on variable incentive or performance fees have increased business risk unless there is a working capital facility in place. Firms with significant assets, where management fees far exceed operating costs, have less business model risk. Larger firms, however, may unfortunately have performance challenges, particularly if they have grown beyond the strategy's capacity or diluted performance just to gain or gather assets.

According to the Merlin study, "There is still no one-size-fits-all business model for hedge funds, but there are several common factors and best practices that have developed to ensure the manager is engaged in a sustainable business. A fund operating in the red zone is dependent on outsized performance to

cover its expenses; a fund in the yellow zone requires minimal performance; and a green zone fund can sustain itself when its performance is lower than expected, nonexistent, or even. negative. Funds that structure their business model to operate in the green zone are better positioned to navigate through downturns and therefore have higher survival rates over the long term."

Figure 10.1 highlights the basic revenue and expense scenarios for three types of hedge fund operating models analyzed in the Merlin report: red zone, yellow zone, and green zone.

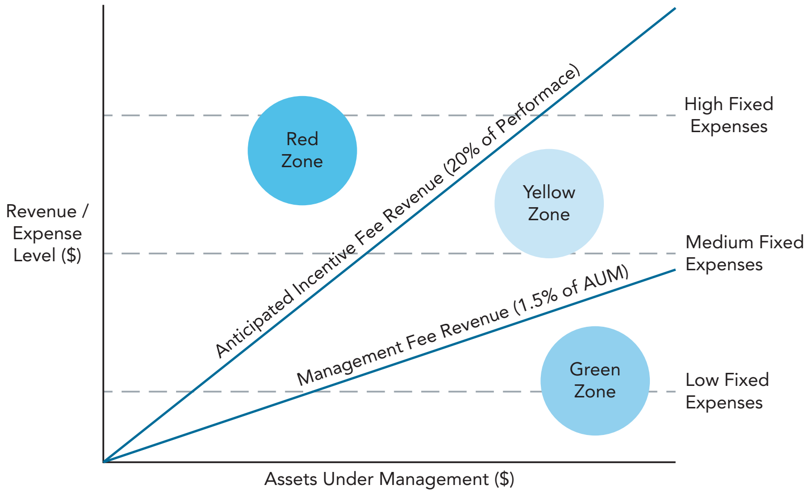


Figure 10.1 **Three types of fund operating models.**

Source: Wells Fargo Securities, "The Business of Running a Hedge Fund Best Practices for Getting to the 'Green Zone,'" February 2011.

The Merlin study noted that it is the performance fee effect that makes the hedge fund model so powerful. Traditional asset management models derive revenues almost exclusively based on assets, whereas a hedge fund's revenues include performance incentives. Figure 10.2 shows the profitability of a hypothetical management company with $3 million in operating costs under a range of performance and asset under management scenarios, some positive and others negative, that was used in the study. Figure 10.3 simply shows the power of incentive or performance fee growth relative to management fees alone.

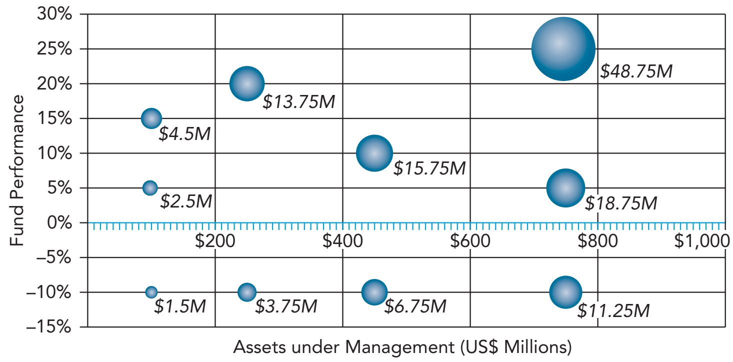


Figure 10.2 **Merlin ADM/Performance MAP.**

Source: Wells Fargo Securities, "The Business of Running a Hedge Fund: Best Practices for Getting to the 'Green Zone,"' February 2011.

In addition to being aware of the firm's strategic positioning, business plan, and other issues that come up regarding its business model, an investor can and should ask questions designed to uncover any additional business model risk. Investors can ask additional questions specifically designed to evaluate a fund's business model risk throughout the entire due diligence process. Questions might include several of the following:

1. What is the firm's strategic vision?
2. Do you have a multiyear budget? Many do not!
3. How many months of cash flow do you have in the bank?
4. What steps can the fund take or is it taking to lower its cost base?
5. When was the last time the fund renegotiated its terms with its key service providers?
6. Does the fund use any outsourced solutions? Why or why not?
7. What is the management company's break-even AUM?
8. What is the fund performance needed to break even at the existing AUM level?
9. What is the capacity of the existing staff to handle additional assets?
10. Does the fund have key man insurance and, if so, on whom, and what is the succession plan for the firm's founder?

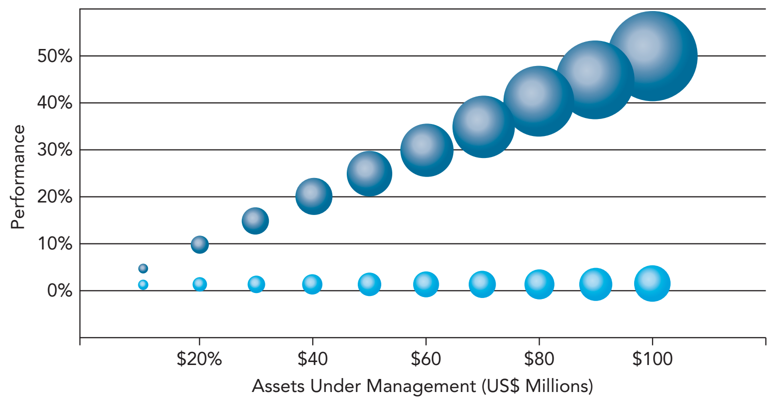


Figure 10.3 **The hedge fund model at work: Performance fee versus management fee growth.**

Source: Wells Fargo Securities, "The Business of Running a Hedge Fund Best Practices for Getting to the 'Green Zone,"' February 2011.

10.10 FRAUD RISK

Investors in hedge funds always need to be on the lookout for fraud. Despite the due diligence done on the investment, risk management, and operational practices of a manager or fund, and even where there is a complete understanding of the business model risk, investors can still find themselves defrauded.

The FBI lists hedge fund fraud as a type of white-collar crime on its website. It says that hedge funds are minimally regulated investments that present many risks to investors. It says that hedge funds can and do fail for many reasons, including leverage or increased risk taking due to negative cash flows to the management company following periods of withdrawals. It says that fraud is another risk present when investing in hedge funds, and it goes on to elaborate on serveral types of frauds that might occur.

According to the FBI website, there are several potential indicators of fraud that investors should investigate before investing in a fund.

- Lack of trading independence when a fund executes via an affiliated broker-dealer.
- Investor complaints about lack of liquidity.
- Litigation in civil court alleging fraudulent acts.
- Unusually strong performance claims.
- A high percentage of illiquid investments or those marked to market by the manager.
- Related parties participating in valuation or a lack of independence (for example, valuation agents, brokers supplying prices, or administrators where the manager is the largest customer, has a personal relationship, or has an investment).
- Personal trading by managers in the same securities or similar securities as the fund.
- Aggressive shorting and organized efforts to spread rumors or disseminate unfounded or materially false information about a company.

The agency suggests that investors review the SEC website for past regulatory actions; state securities websites; and federal district, bankruptcy, and appeals courts records and check service providers' independence and reputation, as well as use of a professional service for background checks. It also encourages anyone who feels that they have been defrauded to report it to the appropriate agency immediately, using the following contacts.

- Securities and Exchange Commission Enforcement: www.sec.gov/enforce
- NASD Investor Complaint Center: www.nasd.com/InvestorInformation/InvestorProtection
- Commodity Futures Trading Commission toll-free complaint line: 866-FON-CFTC
- FBI: https://tips.fbi.gov/

The SEC website also has a section dedicated to hedge fund due diligence. The SEC published a list of questions that investors should consider before making a hedge fund or fund of hedge funds investment. Some of the SEC recommendations include reading a fund's prospectus or offering memorandum and related materials, understanding how a fund's assets are valued, asking questions about fees, being wary of extra layers of fees, understanding limitations on share redemption rights, researching the backgrounds of hedge fund managers, and not being afraid to ask questions.

The SEC has also issued or participated in some important reports on the industry that are useful to review when considering individual managers for investment. The first was issued immediately after the failure of Long-Term Capital Management in 1999, and the second was written in response to the industry's growth in in 2003. "Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management" was written in 1999, and "Implications of the Growth of Hedge Funds" was written in 2003. Another report, prepared for the President's Working Group on Financial Markets by the Asset Management Subcommittee and issued in 2009, outlined recommended best practices for the hedge fund industry in its report appropriately named "Best Practices for the Hedge Fund Industry."

DUE DILIGENCE QUESTIONNAIRE

The table of contents of a typical due diligence questionnaire created by a hedge fund for circulation to potential investors would likely include disclosure of all of the following information.

1. Manager information
   - **a.** Registration
   - **b.** Ownership
   - **c.** Organization
   - **d.** References/Background checks
   - **e.** Track record
   - **f.** Risk management
   - **g.** Operations
   - **h.** Service providers
   - **i.** Contact details
2. Execution and trading
3. Third-party research policy
4. Compliance
5. In-house legal
6. Anti-money-laundering policy and procedures
7. Disaster recovery and business continuity
8. Insurance coverage and key man provisions
9. Fund information
   - **a.** Management and incentive fees
   - **b.** Lock-up
   - **c.** Subscriptions and redemptions
   - **d.** Notice periods
   - **e.** Fund directors or advisors
   - **f.** Administrator
   - **g.** Auditor
   - **h.** Legal advisor
   - **i.** Prime broker
   - **j.** Assets
   - **k.** Performance
   - **l.** Capacity
   - **m.** Gates and lock-ups
   - **n.** Historical drawdown
   - **o.** Use of managed accounts
   - **p.** Investor mix
10. Investment process and portfolio construction
11. Risk controls
    - **a.** Concentration and diversification
    - **b.** Liquidity
    - **c.** Leverage
    - **d.** Exposure to market risk factors
    - **e.** Reporting
    - **f.** Portfolio Greeks (duration, delta, beta, etc.)
12. Financial statements
    - **a.** Year-end opinion
    - **b.** Level 1, 2, 3 assets
    - **c.** Interim statements
    - **d.** Administrator reports
13. Terms and use of third-party marketers

Industry organizations such as COO Connect, a buy-side organization of hedge fund chief operating officers, maintains a number of standard request for proposal documents that can be used to evaluate hedge funds and their service providers, such as fund administrators and prime brokers.

SUMMARY

Performing due diligence on a hedge fund is as much art as it is science. An investor must use a comprehensive checklist to ensure that nothing is left out or omitted yet remain free to ask open-ended questions that provide insights into a firm's philosophy or culture. The guidelines discussed here for investment, operational, and business model risk assessment were for the most part generic. When performing due diligence, investors must modify their approach to be strategy specific. An equity fund requires a different set of questions about its portfolio than a fixed-income fund or a risk arbitrage or distressed fund. An investor in an equity manager might care about the portfolio's beta, whereas an investor in credit is concerned about spreads and rates. Due diligence includes analysis of the investment process, the operational environment, and the fund manager's ability to run a business.

Every fund is unique, and the process is fluid. In the end, an investor needs to balance a set of minimum standards and values with those identified during the due diligence process before deciding to invest. Although there will always be exceptions, it is useful to follow a particular regimen and uniform process when evaluating investments in hedge funds. Doing so ensures a certain amount of efficiency and consistency and ultimately helps investors manage performance and risk expectations over time.

11

Predicting Fraud by Investment Managers

Stephen G. Dimmock, William C. Gerken

Learning Objectives

After completing this reading you should be able to:

- Explain the use and efficacy of information disclosures made by investment advisors in predicting fraud.
- Describe the barriers and costs incurred in implementing fraud prediction methods.
- Discuss how to improve investors' ability to use disclosed data to predict fraud.

Excerpt is reprinted from the Journal of Financial Economics, *Vol 105, Stephen G. Dimmock, William C. Gerken, "Predicting fraud by investment managers," 153–173, 2012, with permission from Elsevier.*

ABSTRACT

We test the predictability of investment fraud using a panel of mandatory disclosures filed with the SEC. We find that disclosures related to past regulatory and legal violations, conflicts of interest, and monitoring have significant power to predict fraud. Avoiding the 5% of firms with the highest ex ante predicted fraud risk would allow an investor to avoid 29% of fraud cases and over 40% of the total dollar losses from fraud. We find no evidence that investors receive compensation for fraud risk through superior performance or lower fees. We examine the barriers to implementing fraud prediction models and suggest changes to the SEC's data access policies that could benefit investors.

11.1 INTRODUCTION

On December 11, 2008, the Securities and Exchange Commission (SEC) charged Bernard Madoff with securities fraud for committing an $18 billion Ponzi scheme.[1] This case emphasized the opportunities advisers have to exploit investors and the importance of limiting advisers' opportunistic behavior through either market or regulatory forces. In the U.S., the regulatory system protects investors primarily through mandatory disclosures. Investment advisers must file Form ADV to disclose information about their operations, conflicts of interest, disciplinary histories, and other material facts. Investors are then responsible for using these disclosures to assess advisers' fraud risk. In this paper, we address the question: Could investors use these mandatory disclosures to predict fraud?

To address this question, we use an annual panel of Form ADVs filed from August 2001 through July 2006. The panel includes 13,853 investment advisers who advise more than 20 million clients and control more than $32 trillion in assets (as of August 2005). These firms advise all mutual funds, nearly all institutional investment funds, and many hedge funds in the U.S. Although the SEC provides public access to each investment adviser's current Form ADV filing, this panel of historical filings is not publicly available, and we are the first researchers to use these data. Our data also include a review of all SEC administrative proceedings and litigation releases from August 2001 through July 2010 to identify those cases in which investment advisers defrauded their clients.

We find that Form ADV disclosures related to past regulatory violations, conflicts of interest, and monitoring are all significant predictors of fraud. Of key importance for investors and regulators, the results show that an investor who avoided the 5% of firms with the highest ex ante predicted fraud risk would avoid 29% of fraud cases and over 40% of the dollar losses from fraud[2] (although to obtain these benefits, the investor would have to forgo investing with 5% of non-fraudulent advisers). Out-of-sample tests confirm the robustness of the fraud predictions.

These findings are subject to several limitations. First, only detected fraud cases are included in the prediction models. Although we conduct extensive out-of-sample tests, we cannot reject the possibility that prediction models are biased because undetected fraud cases are unobservable. Second, although we find that certain characteristics, such as conflicts of interest, can predict fraud, we cannot infer that conflicts of interest cause fraud, or that their prohibition would deter fraud. Prediction does not imply causality, as firms' characteristics may be jointly determined with the decision to commit fraud. Third, in addition to the disclosures mandated by the SEC, investors may assess fraud risk using other sources of information that we do not include in our models. Finally, prediction is not the sole purpose of disclosure; it is also intended to deter fraud. We do not address this deterrent effect of disclosure in this paper.

If the Form ADV data were not useful for predicting fraud, then either disclosure deters fraud so effectively that it eliminates the predictability that would occur in the absence of disclosure or the disclosed information is worthless. Our findings thus provide evidence that regulators require investment advisers to disclose relevant information.

The predictability of fraud raises the question: why do investors allocate money to firms with high fraud risk? One possibility is that the characteristics that predict fraud provide offsetting benefits for investors. For example, affiliation with a brokerage firm could reduce transaction costs or expedite trading. In-house custody of clients' assets could increase fraud risk but reduce costs, resulting in lower fees for investors (e.g., Cassar and Gerakos, 2010). Darby and Karni (1973), Karpoff and Lott (1993), Klein and Leffler (1981), and Lott (1996) argue that if investors differ in their valuation of fraud risk, then some investors would accept a high level of fraud risk in return for superior performance or lower fees, while other investors would choose low fraud risk and accept worse performance or higher fees. To test whether investors receive compensation for fraud risk, we classify investment funds based on their advisers' predicted fraud risk. This subsample includes only the subset of firms that manage funds included in the Trading Advisor Selection System (TASS)

[1] See "SEC Charges Bernard L. Madoff for Multi-Billion Dollar Ponzi Scheme," http://www.sec.gov/news/press/2008/2008–293.htm

[2] For example, the predicted fraud risk of Bernard L. Madoff Investment Securities is above the 95th percentile.

hedge fund, Center for Research in Security Prices (CRSP) mutual fund, and/or Plan Sponsor Network (PSN) Informa databases. For all three types of funds, we find no evidence that investors receive compensation for fraud risk through superior performance or lower fees. However, we cannot rule out the possibility that investors receive some other form of compensation.

Given the surprising result that fraud risk is both predictable and apparently uncompensated, we turn to another possibility. Perhaps barriers to implementing a predictive model cause the costs to outweigh the potential benefits. To explore this possibility, we compare two types of predictive models, both of which take the perspective of an investor attempting to implement a fraud prediction model during the sample period. The first predicts fraud using only the limited subset of information that would have been publicly accessible. Until 2010, the general public had access to only contemporaneous cross-sections of filings; thus, the independent variables in these tests are taken from the contemporaneously accessible filings. The second type of model predicts fraud using a panel of prior filings. These tests show what would have been possible if historical filings have been contemporaneously accessible during the sample period; these models are moderately better at predicting fraud out-of-sample. We discuss simple changes to data access policies that could improve investors' ability to predict fraud.

The paper proceeds as follows. Section 2 discusses the related literature. Section 3 describes our data. Section 4 contains tests of the predictability of fraud. Section 5 tests the relation of fraud risk with the performance and fees of investment funds. Section 6 examines the costs and barriers to predicting fraud. Section 7 concludes.

11.2 RELATED RESEARCH

To our knowledge, just two papers, Bollen and Pool (2010) and Zitzewitz (2006), develop methods to detect fraud by investment advisers. Bollen and Pool (2010) build on earlier studies of hedge funds' manipulation of reported returns (Bollen and Pool, 2008, 2009; Straumann, 2009) and find that suspicious return patterns can predict fraud charges. Zitzewitz (2006) shows that daily fund flows provide information about late trading in mutual funds. Although these papers, like ours, develop methods to detect fraud, they analyze returns and fund flows rather than firms' disclosures of business practices and conflicts of interest. An advantage of using firms' disclosures is that we can actually predict fraud, whereas methods based on returns and flows can only detect past or ongoing fraud. A further advantage is that Form ADV disclosures are mandatory, whereas the disclosure of returns is optional for many investment advisers.

Brown, Goetzmann, Liang, and Schwarz (2008, 2009) examine operational risk using a cross-section of Form ADV filings from hedge fund advisers. The authors define "problem" funds as those managed by an adviser that reports prior legal or regulatory violations, that are either committed by the adviser itself or an affiliated firm. Brown, Goetzmann, Liang, and Schwarz then test whether Form ADV data are associated with prior problems. Because historical Form ADV data are not publicly available, the authors create a measure of operational risk, the w-score, based on the correlations between contemporaneous Form ADV data and historical hedge fund data.

They then test whether the w-score can predict hedge fund closure, flows, and returns. We also use Form ADV data, but our work differs from Brown, Goetzmann, Liang, and Schwarz in several ways. First, we use historical Form ADV filings to make ex ante predictions of fraud. Second, we focus on fraud rather than their very broad definition of operational risk. (Indeed, of the 126 "problem" hedge fund advisers identified by Brown, Goetzmann, Liang, and Schwarz, we find that only six have prior incidents of fraud.) Finally, their measure of operational risk includes violations by affiliated firms, such as broker-dealers. These differences are empirically important; we replicate the w-score of Brown, Goetzmann, Liang, and Schwarz but find it has an insignificant relation with subsequent fraud.

11.3 DATA

Investment Fraud

This study combines two types of data: (1) investment fraud data and (2) disclosures made by investment advisers in their Form ADV filings. To obtain investment fraud data, we search all SEC administrative proceedings and litigation releases[3] that contain the terms "fraud" and "investment adviser" (or "investment advisor") filed from August 2001 through July 2010. From these documents, we identify all cases that involve violations of the antifraud provisions in the Investment Advisers Act. Even when another agency initially detects the fraud case, the SEC launches an administrative proceeding, which we observe. The main dependent variable in our paper includes only fraud cases that harm the firm's investment clients. We do not include insider trading, short sale violations, brokerage fraud, or other crimes, unless they cause direct losses to the firm's investment clients.

Many fraud cases span several years and involve multiple legal actions. Figure 11.1 shows the timeline of a fraud initiated in

[3] See www.sec.gov/litigation/admin.shtml and www.sec.gov/litigation/litreleases.shtml.

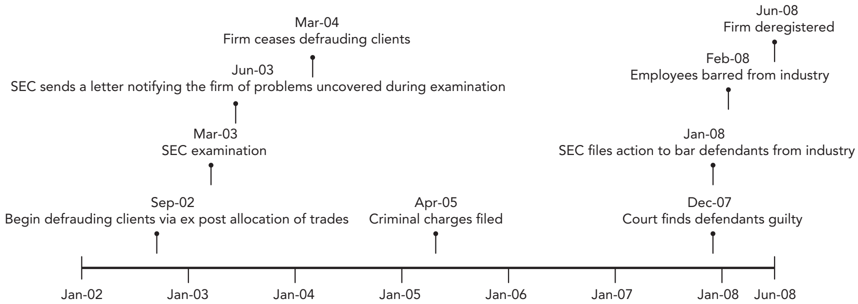


Figure 11.1 One fraud case's timeline.

This figure shows the timeline of one particular fraud, committed by K.W. Brown & Company, from initiation to the end of all legal actions. Beginning in September 2002, the firm began defrauding clients through self-dealing. The firm traded securities for its own proprietary account as well as on behalf of clients. The firm engaged in ex post allocation of trades; securities were purchased but not allocated to a specific account. At a later date, profitable trades were retroactively allocated to the firm's proprietary account and unprofitable tradesv were allocated to clients. This resulted in over \$4.5 million in illegal gains for the firm, and more than \$9 million in client losses.

September 2002 by K.W. Brown & Co., an investment adviser that traded securities on behalf of clients and for its proprietary account. K.W. Brown & Co. purchased securities but delayed assigning them to specific accounts. Eventually, the firm would allocate profitable trades to its proprietary account and unprofitable trades to clients, resulting in losses of over \$9 million to the firm's clients. The SEC uncovered problems in March 2003 during a routine examination and notified the firm in June 2003. The fraud continued until March 2004. In April 2005, the firm and its key employees were charged with fraud. The firm and its employees were convicted in December 2007. In January 2008, the SEC filed an administrative proceeding to bar Kevin W. Brown, his wife, and another employee from the securities industry. The firm was deregistered in June 2008.

Because this kind of extended legal scenario is common, and because our goal is to predict fraud rather than detect it, we aggregate all legal actions associated with a single underlying fraud into a single "case" and identify the periods in which fraud occurred. For example, we define the K.W. Brown & Co. fraud case as occurring from September 2002 until March 2004, and use the information from K.W. Brown & Co.'s August 2002 Form ADV filing to predict the initiation of fraud in September 2002. We also use information from K.W. Brown & Co.'s August 2003 Form ADV filing to predict the continuation of fraud into 2004. For the remaining years of the sample, we classify K.W. Brown & Co. as a clean firm. By predicting the occurrence of fraud in 2002–2004, rather than its detection in 2005, we avoid potential biases caused by a correlation between detection and time variation in the predictive variables.

The SEC legal filings include investment fraud cases committed by firms that did not register with the SEC, and thus were not required to file Form ADV (see the next subsection for more detail). To address the economic importance of fraud committed by registered versus non-registered investment advisers, Panel A of Table 11.1 summarizes fraud cases for both types of firms. Registered firms commit slightly over half of investment fraud cases and are responsible for the overwhelming majority of the dollar losses from fraud. Thus, although the scope of our tests is limited to registered investment advisers, these firms are responsible for the most economically meaningful fraud cases.

Panel B summarizes firm-wide fraud, committed with the knowledge of the firm's executive officers, as well as fraud by rogue employees who evade their firms' internal controls. The vast majority of fraud cases are firm-wide. Panel B also summarizes the dollar losses and the duration of the fraud cases. Because fraud often involves the falsification of records, some loss amounts are unavailable, and the available amounts are generally a lower bound, including only the proven losses. Fraud duration is defined as the period extending from the initiation of the fraud until the firm ceases the fraudulent activity. The median fraud case persists for nearly three years. The maximum durations, summarized in the last column of Panel B, reflect the fact that the sample includes cases that were initiated prior to 2001.

To test whether past fraud can predict future fraud, we search SEC administrative proceedings and litigation releases filed from September 1995[4] through July 2001, and create two variables. The first, Past fraud, is equal to one if a prior administrative proceeding or litigation release shows that the firm has committed fraud. The second, Past affiliated fraud, is equal to one if an

[4] Online access to administrative proceedings and litigation releases begins in September 1995.

Table 11.1 **Summary of Investment Frauds**

This table summarizes cases of investment fraud committed by investment advisers between August 2001 and July 2010 as reported in SEC administrative proceedings and litigation releases. Registered denotes firms that file a Form ADV with the SEC. Firm-wide fraud is committed by high level executives, or at the very least, with the firms' implicit acceptance. Rogue employee fraud is committed by individuals who evade their firms' internal control systems and the firms do not knowingly benefit.

Panel A: Registered versus Non-Registered Advisers				
	Total	**Firm-Wide**	**Rogue Employee**	**Total Sum ($ billion)**
Non-Registered	251	244	7	4.5
Registered	258	217	41	32.4
Total	509	461	48	36.9

Panel B: Fraud Characteristics								
		Investor Losses ($ million)				**Duration (years)**		
Classification	**Obs.**	**Mean**	**Median**	**Max**	**Missing Obs.**	**Mean**	**Median**	**Max**
Firm-wide	217	196.3	6.0	18,000.0	56	4.0	3.1	20.8
Rogue employee	41	25.4	3.0	300.0	8	3.2	2.5	11.1
Total	258	167.2	5.1	18,000.0	64	3.9	2.9	20.8

affiliated firm has committed fraud (affiliation implies the firms are under common control, such as common ownership or executives). Both variables are restricted to include only fraud cases that harmed investment advisory clients. This restriction is consistent with the main dependent variable. Further, Karpoff and Lott (1993), Karpoff, Lott, and Wehrly (2005), and Murphy, Shrieves, and Tibbs (2009) show that firms suffer greater reputational penalties for defrauding counterparties, such as customers, than for defrauding other stakeholders. We match Past affiliated fraud to investment advisers using the affiliated firm identifiers from Schedule D of Form ADV. To prevent a look-ahead bias in the predictive regressions, these variables only include fraud cases that had ceased and were publicly revealed before August 1 of that year. For example, in the K.W. Brown & Co. case summarized in Figure 11.1, Past fraud is equal to one only after April 2005 when the fraud had ceased and the first relevant SEC legal filing was publicly accessible.

Form ADV Data

The Investment Advisers Act, which expressly defines and prohibits investment adviser fraud, requires all advisers with more than $25 million in assets under management and with 15 or more U.S. clients to register with the SEC. The Act defines an investment adviser as any entity that receives compensation for managing securities portfolios or providing advice regarding individual securities.[5] Registered investment advisers must file Form ADV to disclose past regulatory violations and potential conflicts of interest.

Form ADV contains 12 items and four schedules. Items 1–6 contain descriptive information about a firm and its operations. Items 7 and 8 require disclosure of certain conflicts of interest. Item 9 requires disclosure regarding the custody of clients' assets. Item 10 requires disclosure of control persons. Item 11 requires disclosure of past legal and regulatory violations. Item 12 identifies small businesses. Schedules A–C identify the direct and indirect owners of a firm. Schedule D requires disclosure of affiliations with other financial firms.

An SEC Web site provides a public link to the Investment Adviser Registration Depository, which includes the most recent Form ADV filings from all registered investment advisers.[6] Until recently, investors could access the latest filings only one at a time, and

[5] Section 203(b)(3) of the Investment Advisers Act exempts firms with fewer than 15 U.S. clients, that do not advise funds registered under the Investment Company Act, nor "hold themselves out to the public" as investment advisers. Some hedge funds use this exemption to avoid registration. A 2004 SEC ruling required hedge fund advisers to register by February 2006, but a U.S. District Court reversed this ruling in June 2006. Despite these exemptions, many hedge fund advisers were registered prior to 2006, either voluntarily or because they also advised other portfolios.

[6] See www.sec.gov/IARD

past filings were unavailable. Beginning in January 2010, the SEC began to provide downloadable files of historical Form ADV data.[7] Downloadable files from July 2006 through November 2009 contain summaries of the schedules rather than Form ADV's item data. Downloadable files from December 2009 until the present contain the item data, but not the schedule data.

The SEC provided us with a database of all Form ADV filings from August 2001 through July 2006, including initial filings, amendments, schedules, and the filings of now-defunct firms. These data are not publicly accessible and, to our knowledge, no other researchers have examined them. To create an annual panel for the predictive regressions, we select each firm's most recent filing as of August 1 of each year.[8] This annual panel includes 53,994 firm-year observations representing 13,853 unique investment management firms. We combine the investment fraud documentation and Form ADV data by matching the firms' full legal names.[9]

Form ADV Variables

Table 11.2 summarizes a cross-section of the investment advisers' characteristics and disclosures, using information from each firm's first Form ADV filing during the sample. Panel A shows that the median firm is wholly employee-owned. Employee ownership, calculated as in Dimmock, Gerken, and Marietta-Westberg (2011), is included because external owners may deter fraud by monitoring employees. The Average account size is $55 million, but this variable is highly skewed and the median is only $1.4 million. Percent client agents is the percentage of the firm's clients who are agents (e.g., pension fund managers) rather than the direct beneficiaries of the invested funds. On average, 23.2% of a firm's clients are agents. This additional layer of agency is potentially related to fraud because agents have weaker incentives to monitor investment advisers, but may also have greater expertise and financial sophistication. Assets under management (AUM) varies greatly across firms. The median AUM is $90 million, but the mean is greater than $2.2 billion.

Panel B of Table 11.2 tabulates many of the variables disclosed in Form ADV (see the Appendix for detailed definitions). Column 1 shows summary statistics for the full sample. Column 2 shows summary statistics for firms in which no fraud is committed from August 1, 2001 through July 2007 (Clean). Column 3 shows summary statistics for firms in which fraud is committed during the sample period (Fraud). The third column also reports the univariate significance of the difference between clean and fraud firms, using Fisher's exact test.

Item 11 of Form ADV requires each investment adviser to disclose its disciplinary history, as well as that of its (non-clerical) employees, its affiliated firms, and the employees of affiliated firms. The 24 questions in Item 11 are divided into three categories: regulatory, criminal, and civil judicial. From these questions, we create two indicator variables. Past regulatory equals one if the firm discloses past regulatory violations, indicating sanctions by the SEC, the Commodity Futures Trading Commission, or a self-regulatory organization such as the Financial Industry Regulatory Authority (FINRA). The second variable combines the remaining two categories; Past civil or criminal equals one if the firm discloses unfavorable civil judicial decisions related to investment advising, or if the firm discloses criminal convictions. Fraud firms are significantly more likely to report both types of violations.

The disclosure information in Item 11 covers a wide range of regulatory and legal offenses, and the offenses are often minor, such as failing to follow protocols for record storage. Minor violations seem to be the norm rather than the exception, and should be interpreted as such: Less than 2.5% of firms that report past violations have a prior instance of fraud. Form ADV does not distinguish whether the investment adviser or its affiliate(s) committed the reported violations, and so there is a strong positive correlation between prior violations and the number of affiliates. To avoid a spurious correlation between the prior violations of affiliated firms and investment adviser fraud, our dependent variables do not include fraud committed by affiliated firms.

Items 7 and 8 of Form ADV require firms to disclose conflicts of interest. From this information we create three variables. Referral fees equals one if the firm compensates other parties for client referrals. Interest in transaction equals one if the firm trades directly with its clients or has a direct financial interest in securities recommended to its clients; these practices create potential conflicts and provide a mechanism for fraud. Soft dollars equals one if the firm directs clients' trades to a brokerage with relatively high commissions and, in return, the broker supplies the adviser with research or other benefits. Since clients pay the costs while the investment adviser realizes the benefits, soft dollars create a potential conflict of interest.

The next four variables are intended to measure monitoring. Broker in firm equals one if the firm employs registered representatives of a broker-dealer. Trading through an affiliated broker-dealer removes one form of external oversight and provides a

[7] See www.sec.gov/foia/docs/invafoia.htm

[8] Firms must file Form ADV at least once per year, but often file more frequently; the median firm files 11 times per year. We choose August 1 to maximize the number of annual observations since our set of Form ADV filings ends July 31, 2006.

[9] Of the 251 fraud cases committed by non-registered investment advisers, 13 were registered at the time the fraud was initiated, but were deregistered before our sample began. We successfully match all fraud cases by currently registered investment advisers.

Table 11.2 **Summary of Investment Advisory Firms**

This table summarizes information from each firm's first Form ADV filing during the period August 2001 through July 2006. There are 13,853 unique firms in the sample. Employee ownership is the aggregate employee ownership of the firm. Percent client agents is the percentage of clients that are agents for the owners of the assets. Past fraud equals one if the firm is identified as committing fraud in a previous SEC filing. Past affiliated fraud equals one if the firm's affiliates have been identified as committing fraud in a previous SEC filing. Past regulatory equals one if the firm reports past regulatory violations. Past civil or criminal equals one if the firm reports past civil or criminal violations. Referral fees equals one if the firm compensates any party for client referrals. Interest in transaction equals one if the firm recommends securities in which it has an ownership interest, serves as an underwriter, or has any other sales interest. Soft dollars equals one if the firm receives benefits other than execution from a broker-dealer in connection with clients' trades. Broker in firm equals one if the firm employs registered representatives of a broker-dealer. Investment Company Act equals one if the firm is registered under the Investment Company Act of 1940. Custody equals one if the firm has custody of clients' cash or securities. Dedicated CCO equals one if the chief compliance officer has no other job title. Hedge fund clients equals one if more than 75% of the firm's clients are hedge funds. The column Clean (Fraud) summarizes firms in which a fraud is not committed (is committed) from first filing through July 2007. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels based on Fisher's exact test.

Panel A: Firm Characteristics					
	Mean	**SD**	**25th**	**50th**	**75th**
Employee ownership	68.2%	44.2	0.0	100.0	100.0
Avg. acct. size ($ thousand)	55,361	328,522	339	1,442	21,667
Percent client agents	23.2%	32.6	0.0	8.3	30.0
Assets under mgmt. ($ million)	2,213	16,433	37	90	400
Firm age (years)	5.1	7.7	0.4	1.1	8.1
Panel B: Firm Disclosures					
			All	**Clean**	**Fraud**
Past fraud			0.2%	0.2	1.6***
Past affiliated fraud			1.6%	1.6	2.6
Past regulatory			12.1%	11.9	32.6***
Past civil or criminal			3.3%	3.1	12.5***
Referral fees			40.0%	39.7	59.8***
Interest in transaction			30.4%	30.1	52.2***
Soft dollars			55.7%	55.6	63.0**
Broker in firm			40.8%	40.4	66.3***
Investment Company Act			9.8%	9.6	29.0***
Custody			23.9%	23.7	33.7***
Dedicated CCO			10.7%	10.7	12.4
Hedge fund clients			13.4%	13.5	6.2***

mechanism for fraud. Investment Company Act equals one if the firm manages money on behalf of a fund registered under the Investment Company Act, such as a mutual fund. The Act increases regulation and disallows certain conflicts of interest but also indicates the firm's investors are relatively unsophisticated. Custody equals one if the firm has possession, or the authority to obtain possession, of its clients' assets. Custody facilitates fraud by removing external oversight. However, SEC Rule 206(4)–2 requires audits of investment advisers with such custody, including at least one unannounced visit per year, which may reduce the incentive for fraud by increasing the likelihood of detection. Dedicated CCO equals one if the firm's chief compliance officer (CCO) does not have another formal job title. All registered investment firms must designate a CCO who is responsible for ensuring compliance with SEC regulation, but often the CCO has other potentially conflicting roles within the firm.

Hedge fund clients equals one if over 75% of the firm's clients are hedge fund clients. We include this variable for two reasons: First, hedge funds are relatively opaque, which could facilitate fraud. Second, prior to 2006, some hedge fund advisers were not required to file Form ADV, which could create a sample selection bias if non-reporting is associated with fraud.

Fund-Level Data: Returns and Fees

To test the relation of fraud risk with performance and fees, we require fund-level data that are not disclosed in Form ADV. We obtain fund-level data from the TASS hedge fund, CRSP mutual fund, and PSN Informa databases. We match these databases to the Form ADV sample using firm name, location, and assets under management.[10] For the CRSP mutual fund and PSN Informa databases, we include only equity funds in our sample.

We are able to match 1,511 of the firms in the TASS database (37.2%), which manage 2,848 distinct hedge funds. From TASS we obtain monthly returns, management fees, incentive fees, and other variables. Participation in the TASS database is voluntary and hedge funds are not required to publicly disclose their returns. As a result, the merged hedge funds may not be representative of all hedge funds.

Because all mutual fund advisers must file Form ADV and publicly report their returns, we are able to match all management companies in the CRSP mutual fund database. From this database we obtain monthly returns, expense ratios, and other variables for 2,818 equity funds.

To obtain information on institutional funds, we use the PSN Informa database [institutional funds are long-only portfolios managed on behalf of accredited investors; see Busse, Goyal, and Wahal, 2010 for more details]. We are able to match 1,578 of the PSN firms (88.2%), which manage 4,189 distinct portfolios and 89.2% of the aggregate assets under management. Like hedge funds, institutional funds are not required to publicly disclose their returns, and so the funds in our sample may not be representative. In addition to monthly returns, we also obtain information on the posted annual fee charged on a $50 million account (institutional funds can charge clients different fees and the reported fees are only approximate).

In total, 3,123 of the firms in the Form ADV sample match to at least one of the fund return databases, and 314 of the firms match to all three databases. Although the matched firms are only 22.5% of the total Form ADV sample, they control the majority of assets under management.

11.4 PREDICTING FRAUD

In this section, we test whether the Form ADV data can predict investment fraud. The purpose of these tests is prediction and, as noted previously, we make no claims regarding causality. Many of the independent variables are endogenous (e.g., a firm's executives may deliberately choose an organizational structure that enables fraud), but because our goal is prediction rather than establishing causality, the potential endogeneity of the independent variables does not change our interpretation.

A major caveat in interpreting our findings is that we observe only detected fraud. Three factors affect observed fraud: the unobservable true rate of fraud, the probability of detection given a fixed level of monitoring, and the allocation of monitoring resources. Ideally, the regressions will predict the true rate of fraud. However, if certain predictive variables are correlated with either monitoring or detection, this relation could affect the interpretation of the results. Further, the predictive variables could be correlated with monitoring and detection for two reasons. First, any predictive variable that decreases the probability of detection will increase the incentive to commit fraud. In general, this problem biases against significant results because predictive variables that are associated with a higher rate of fraud will also be associated with a lower detection rate. Second, if the difficulty of detecting fraud affects the allocation of monitoring resources, this may, or may not, outweigh the added difficulty of detecting fraud. These difficulties could cause the empirically observed relations to differ from the actual relation between firms' disclosures and the unobservable true rate of fraud.

[10] For summary statistics and details about matching, see the Web Appendix.

We address the issue of undetected fraud in two ways. First, we conduct extensive out-of-sample tests to ensure the predictions are robust. Second, although the panel of independent variables ends in 2006, we search for detected fraud cases through July 2010. For each case, we identify when the fraud occurred. In the predictive regressions, the dependent variable is the occurrence of fraud in a given year, even if the fraud remains undetected for years. Unfortunately, a direct test of the relation between these variables and fraud detection is not possible. Certain types of fraud may go undetected, a possibility that could bias the results.

Prediction Models

Panel A of Table 11.3 shows the results of probit regressions that predict investment fraud using Form ADV disclosures. In column 1, the sample is a cross-section of firms. The independent variables are taken from each firm's first Form ADV filing during the sample period; the dependent variable equals one if the firm commits fraud at any time between its first filing during the sample period and July 2007. This specification includes indicator variables for the year in which the firm first filed Form ADV. In this cross-sectional specification, the Z-scores are based on robust standard errors.

In the remaining columns, the sample is an unbalanced panel of firm-year observations. The dependent variable equals one if a fraud occurs during the subsequent 12 months. In columns 2 and 3, the sample includes all firm-year observations. In column 4, the sample excludes firms with a history of fraud identified in prior SEC administrative proceeding or litigation releases. In the last column, the sample also excludes firms that disclose the relatively minor legal or regulatory violations in Item 11, committed by either the firm itself or an affiliated firm. For these panel specifications in columns 2–5, the *Z*-scores are based on standard errors clustered by firm and year. The chi-square tests at the bottom of each column show the significance of the overall model.

Past fraud is insignificant in both the cross-sectional and panel regression. There are few firm-year observations with prior fraud because many firms that commit fraud subsequently cease operations, and so the regressions have low power with respect to this variable.[11] Past affiliated fraud does not predict fraud in the cross-sectional regression, but in one of the panel regressions, the coefficient is marginally significant and negative. Unlike the other predictive variables, Past fraud and Past affiliated fraud are not disclosed in Form ADV.

Past regulatory and Past civil or criminal are both significant positive predictors of subsequent fraud, even in the sample that

[11] See Web Appendix Table E for information on firm survival following fraud.

Table 11.3 Predicting Fraud

The full sample consists of 53,994 firm-year observations. In the first column, the sample includes only each firm's first Form ADV filed during the sample period. In the remaining columns, the independent variables are taken from each firm's Form ADV filing as of August 1 each year from 2001 through 2006. In the second and third columns, the full sample is included. In the fourth column, the sample excludes firms with a previously disclosed fraud. In the fifth column, the sample excludes all firms that disclose in Item 11 of Form ADV any type of prior legal or regulatory violation, either by the firm itself or an affiliated firm. Refer to Table 11.2 for variable definitions. Column 1 of Panel A shows the results of a cross-sectional probit regression predicting fraud. The dependent variable equals one if the firm commits fraud in any subsequent year through July 2007. Standard errors are robust. Columns 2–5 show the results of pooled probit regressions predicting fraud. The dependent variable equals one if the firm commits fraud in the subsequent year. Standard errors are clustered by firm and year. In the interest of brevity, the constants are not reported. *Z*-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The columns in Panels B, C, and D correspond to the columns in Panel A. Panel B shows the proportion of fraud that could be predicted within-sample. Panel C shows the out-of-sample performance of each model, using Form ADV filings in August 2006 to predict fraud cases that occur between August 2007 through July 2010. Panel D shows the results from K-fold cross-validation tests.

Panel A: Predictors of Fraud					
	Cross Section	**Full Sample**	**Full Sample**	**No Prior Fraud**	**No Prior Violations**
Past fraud	0.329 [0.98]		0.272 [1.46]		
Past affiliated fraud	−0.224 [1.14]		−0.184 [1.54]	−0.196* [1.68]	

00011693-00000002_CH11_p201-228.indd 209 05/07/22 3:08 PM

Table 11.3 Continued

Panel A: Predictors of Fraud					
	Cross Section	Full Sample	Full Sample	No Prior Fraud	No Prior Violations
Past regulatory	0.175**	0.284***	0.282***	0.285***	
	[2.25]	[4.20]	[4.16]	[4.15]	
Past civil or criminal	0.223*	0.191**	0.200**	0.209**	
	[1.93]	[2.13]	[2.29]	[2.32]	
Referral fees	0.135**	0.100*	0.099*	0.099*	0.139**
	[2.09]	[1.79]	[1.79]	[1.78]	[2.40]
Interest in transaction	0.138*	0.197***	0.198***	0.196***	0.184**
	[1.93]	[2.89]	[2.91]	[2.86]	[2.24]
Soft dollars	−0.029	−0.051	−0.046	−0.040	−0.073
	[0.46]	[0.89]	[0.81]	[0.71]	[1.10]
Broker in firm	0.237***	0.118**	0.120**	0.120**	0.096
	[3.89]	[2.01]	[2.05]	[2.02]	[1.55]
Investment Co. Act	0.103	0.263***	0.269***	0.278***	0.273***
	[1.39]	[3.29]	[3.36]	[3.58]	[2.83]
Custody	0.309***	0.094	0.097	0.088	0.028
	[3.79]	[1.43]	[1.50]	[1.36]	[0.33]
Dedicated CCO	0.288***	−0.088	−0.085	−0.085	−0.056
	[2.67]	[0.86]	[0.82]	[0.82]	[0.53]
Majority emp. owned	0.045	0.009	0.001	0.008	0.033
	[0.65]	[0.11]	[0.02]	[0.10]	[0.37]
Log (avg. acct. aize)	−0.043***	−0.072***	−0.070***	−0.065***	−0.028
	[3.45]	[4.25]	[4.12]	[3.68]	[1.12]
Percent client agents	0.001	0.003***	0.003***	0.003***	0.003***
	[1.40]	[3.91]	[3.88]	[3.77]	[2.91]
Hedge fund clients	−0.035	0.031	0.031	0.020	0.030
	[0.27]	[0.27]	[0.27]	[0.18]	[0.22]
Log (AUM)	0.036***	0.060***	0.059***	0.054***	0.020
	[3.76]	[4.10]	[3.98]	[3.57]	[0.93]
Log (firm age)	0.014	0.002	0.002	0.002	0.008
	[1.18]	[0.20]	[0.19]	[0.20]	[0.66]

Panel A: Predictors of Fraud					
	Cross Section	**Full Sample**	**Full Sample**	**No Prior Fraud**	**No Prior Violations**
Model chi-square	175.2***	181.5***	198.9***	176.9***	63.2***
Observations	13,853	53,994	53,994	53,750	45,920
Panel B: Within-Sample Predictions					
# Fraud	193	517	517	501	310
Fraud predicted	59	150	152	140	44
	30.6%	29.0	29.4	27.9	14.2
# Clean firms	13,660	53,477	53,477	53,249	45,610
Clean firm false positives	683	2,673	2,673	2,662	2,280
	5.0%	5.0	5.0	5.0	5.0
Prop. of $ losses avoided	37.4%	41.3	43.0	40.5	7.9
Panel C: Out-of-Sample Predictions (August 2007–July 2010)					
	Cross Section	**Full Sample**	**Full Sample**	**No Prior Fraud**	**No Prior Violations**
# Fraud	27	27	27	25	18
Fraud predicted	9	10	9	7	1
	33.3%	37.0	33.3	28.0	5.6
# Clean firms	10,356	10,356	10,356	10,293	8,912
Clean firm false positives	517	517	517	514	445
	5.0%	5.0	5.0	5.0	5.0
Panel D: K-Fold Cross-Validation Hold-Out Sample Predictions (August 2001–July 2007)					
Avg # fraud predicted	53.6	143.3	142.4	129.7	35.0
Avg % fraud predicted	27.8%	27.7	27.5	25.9	11.3
Stdev fraud predicted (#)	1.39	3.64	3.75	4.32	2.66
Min # fraud predicted	51	135	133	120	32
Max # fraud predicted	56	149	148	137	42
Avg # false positives	678.4	2,669.2	2,669.2	2,658.2	2,275.8
Avg % false positives	5.0%	5.0	5.0	5.0	5.0
Stdev false positives	0.68	0.95	0.95	0.99	0.91
Min # false positives	677	2,668	2,668	2,656	2,274
Max # false positives	679	2,671	2,671	2,660	2,277

excludes firms with prior fraud. The simplest explanation is that past problems, although frequently minor, indicate poor internal controls or unethical management. But two additional explanations exist: Past violations could increase the rate of detected fraud due to the increased probability of an SEC examination. Also, because each firm must disclose both its own prior violations and those of its affiliated firms, prior violations are strongly correlated with the size and scope of an investment firm's affiliated businesses (i.e., financial conglomerates are more likely to report prior violations). These affiliations could increase conflicts of interest and provide the means to commit fraud.

The next three variables measure several potential conflicts of interest between investment advisers and their clients. Referral fees has a significant positive relation with subsequent fraud. Fraud firms could be relatively willing to pay referral fees because fraud increases the marginal profit per dollar managed. Interest in transaction also has a significant positive relation with subsequent fraud. When investment managers take the opposite side of a transaction from their clients, this arrangement creates an obvious conflict of interest and also provides a mechanism for fraud. Soft dollars does not significantly predict fraud.

We include several variables to measure the monitoring of investment advisers. Broker in firm has a significant positive relation with subsequent fraud. Trading through an in-house brokerage removes external oversight and creates a mechanism for committing certain types of fraud. Investment Company Act has a significant positive relation with subsequent fraud. The Act increases regulatory oversight of these firms, which could increase the probability that fraud is detected. Alternatively, the true rate of fraud could be higher because these firms exploit their clients' lack of financial sophistication. The next three variables, Custody, Dedicated CCO, and Majority employee owned, are not significant in the panel regressions, although Custody and Dedicated CCO are significant in the cross-sectional regression. Note that even if a variable is insignificant in these regressions, its disclosure may still have a beneficial deterrence effect.

The next three variables (Logarithm of average account size, Percent client agents, and Hedge fund clients) also measure monitoring but are based on client characteristics. Although all clients have an incentive to monitor, large investors have a stronger incentive and possibly a greater ability to do so. The results for the Logarithm of average account size show that larger investors are associated with fewer subsequent fraud cases. This result could be a selection effect, meaning that large investors select honest managers. Alternatively, because of financial sophistication or economies of scale in monitoring, large investors may deter fraud because of a higher probability of detection. Both arguments suggest that large investors are associated with a lower rate of fraud rather than a lower detection rate.

Percent client agents, the second variable measuring client characteristics, has a significant positive relation with subsequent fraud. After conditioning on average account size, firms whose clients include a high proportion of agents are more likely to commit fraud. Although agents may have reputational concerns and greater financial sophistication, they do not bear the full cost of fraud, which reduces their incentive to monitor and suggests that they can be swayed through gifts or kickbacks. The reduced incentives of agents appear to outweigh their potentially higher sophistication.

Hedge fund clients is an indicator for firms that primarily manage hedge funds. The results do not provide evidence of a relation between hedge fund management and fraud. Hedge funds are relatively non-transparent, however, and so the detected fraud cases may understate the true frequency of fraud that occurs within hedge funds. Moreover, not all hedge funds were required to file Form ADV during the early part of the sample, which could create a sample selection bias. Nonetheless, in annual cross-sectional regressions (Table 11.4), we find that the coefficient on hedge fund management is not significantly different in the later years of the sample, which suggests that sample selection is not a problem.

The Economic Interpretation of the Prediction Models

The probit regressions in Panel A of Table 11.3 show that the Form ADV variables have a statistically significant relation with subsequent fraud. This finding is important, but the key question of interest is whether the overall model would enable an investor to avoid fraud. To address this question, we take the predicted values from the regressions and examine the tradeoff between correctly predicted fraud cases and the false positive rate. False positives, which occur when the model incorrectly predicts that a clean firm will commit fraud in the subsequent year, can be interpreted as the opportunity cost to investors of erroneously limiting their investment opportunity set. Although failing to predict fraud is likely more costly than mistakenly avoiding an honest investment adviser, an investor would need to avoid multiple honest advisers for every fraud avoided. We address the costs of false positives in Section 5.

To illustrate the possible tradeoffs between false positives and predicted fraud, Figure 11.2 shows a receiver operating characteristic (ROC) curve for the prediction model in the second column of Panel A of Table 11.3. The points on the ROC curve are generated non-parametrically by taking each observation's

Table 11.4 **Annual Cross-Sectional Regressions**

The sample consists of 53,994 firm-year observations. Each column contains an annual cross-sectional regression, in which the independent variables are taken from each firm's Form ADV filing as of August 1 each year from 2001 through 2006. Panel A shows the results of annual cross-sectional probit regressions predicting fraud. The dependent variable equals one if the firm commits fraud in the subsequent year. Refer to Table 11.2 for variable definitions. In the interest of brevity, we do not report coefficients for the constants. Standard errors are robust. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The columns in Panel B correspond to the columns in Panel A. Panel B shows the proportion of fraud that could be predicted within-sample.

Panel A: Predictors of Fraud						
	Aug '01	**Aug '02**	**Aug '03**	**Aug '04**	**Aug '05**	**Aug '06**
Past fraud	0.280	0.172	−0.337	0.451	0.625**	0.553*
	[0.82]	[0.52]	[0.78]	[1.49]	[2.25]	[1.81]
Past affiliated fraud	−0.321	−0.241	−0.194	−0.010	0.201	−0.051
	[1.23]	[1.15]	[1.11]	[0.05]	[0.94]	[0.19]
Past regulatory	0.187*	0.235**	0.350***	0.371***	0.248**	0.050
	[1.89]	[2.42]	[3.79]	[3.44]	[2.01]	[0.37]
Past civil or criminal	0.239*	0.212*	0.092	0.328**	−0.095	0.342*
	[1.73]	[1.66]	[0.71]	[2.41]	[0.47]	[1.76]
Referral fees	0.041	0.021	0.071	0.130	0.187*	0.255**
	[0.45]	[0.26]	[0.85]	[1.51]	[1.90]	[2.25]
Interest in transaction	0.265***	0.252***	0.235**	0.138	0.124	−0.029
	[2.78]	[2.63]	[2.48]	[1.23]	[1.03]	[0.23]
Soft dollars	−0.037	−0.067	−0.015	−0.020	−0.075	−0.190
	[0.40]	[0.78]	[0.17]	[0.21]	[0.74]	[1.53]
Broker in firm	0.202**	0.127	0.071	0.043	0.006	0.147
	[2.33]	[1.51]	[0.84]	[0.45]	[0.05]	[1.23]
Investment Co. Act	0.245**	0.325***	0.306***	0.264**	−0.120	−0.081
	[2.43]	[3.25]	[2.89]	[2.26]	[0.73]	[0.42]
Custody	0.006	0.084	0.166*	0.061	0.246**	0.339***
	[0.06]	[0.89]	[1.76]	[0.58]	[2.37]	[2.65]
Dedicated CCO	0.247	0.348**	0.420***	−0.083	0.057	−0.126
	[1.53]	[2.54]	[3.41]	[0.59]	[0.58]	[1.08]
Majority emp. owned	−0.089	−0.110	0.026	0.220**	0.252**	−0.022
	[0.88]	[1.14]	[0.28]	[2.26]	[2.23]	[0.17]

Table 11.4 Continued

Panel A: Predictors of Fraud						
	Aug '01	**Aug '02**	**Aug '03**	**Aug '04**	**Aug '05**	**Aug '06**
Log (avg. acct. size)	−0.100***	−0.089***	−0.076***	−0.056**	−0.034	−0.061
	[3.90]	[4.18]	[3.29]	[2.14]	[1.02]	[1.61]
Percent client agents	0.004***	0.003**	0.004***	0.002*	0.003*	0.001
	[3.03]	[2.57]	[2.98]	[1.74]	[1.72]	[0.50]
Hedge fund clients	0.006	0.098	0.150	0.127	−0.091	−0.070
	[0.02]	[0.46]	[0.79]	[0.61]	[0.44]	[0.28]
Log (AUM)	0.091***	0.080***	0.062***	0.041*	0.028	0.046
	[4.24]	[4.49]	[3.21]	[1.94]	[1.08]	[1.62]
Log (firm age)	0.019	0.007	−0.003	0.009	0.017	0.000
	[1.00]	[0.24]	[0.13]	[0.36]	[0.72]	[0.00]
Model chi-square	138.0***	138.7***	135.3***	78.1***	44.6***	63.9***
Observations	7,352	7,747	8,562	9,088	10,862	10,383
Panel B: Within-Sample Predictions						
# Fraud	104	116	115	83	59	40
Fraud predicted	39	45	37	22	11	10
	37.5%	38.8	32.2	26.5	18.6	25.0
Clean firms	7,248	7,631	8,447	9,005	10,803	10,343
Clean firm false positives	362	381	422	450	540	517
	5.0%	5.0	5.0	5.0	5.0	5.0

predicted value from the probit model as a cut-point, and then computing both the proportion of fraud firm-years correctly predicted and the false positives. Random prediction of fraud would result in a straight 45-degree line. Initially, the curve rises steeply, showing a considerable number of fraud firm-years could be avoided at a low false positive rate.

The ROC curve in Figure 11.2 shows the full range of all possible tradeoffs between the prediction of fraud and false positives. Following a similar format as Dechow, Ge, Larson, and Sloan (2011), in Panel B of Table 11.3 we provide greater detail for one possible tradeoff, the proportion of fraud firm-years that could be predicted within-sample at a false positive rate of 5%. The columns in Panel B correspond to the columns in Panel A. For example, the model in the second column correctly predicts 150 of 517 fraud firm-years (29.0%) at a false positive rate of 5% (we incorrectly predict fraud in 2,673 clean firm-years that are associated with 885 distinct firms). The last row of Panel B shows the percentage of total dollar losses that could have been avoided at a false positive rate of 5%. The dollar losses from fraud are winsorized at the 99th percentile because of several extreme outliers (e.g., Madoff's $18 billion Ponzi scheme). For multiyear fraud cases, we distribute the losses evenly across years. The model in the second column correctly predicts 41.3% of the total dollar losses from fraud at a false positive rate of 5%, which indicates that the model predicts economically meaningful fraud cases and not merely small cases.

The results in Panel B are similar for all models, except for the specification reported in the last column, in which the subsample does not include firms that report prior legal or regulatory violations, either by the firm or its affiliates. For this subsample,

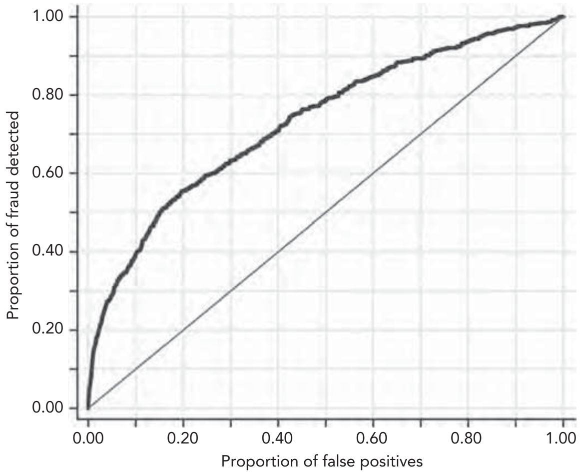


Figure 11.2 Proportion of fraud predicted for all false positive rates.

This figure shows the receiver operating characteristic (ROC) curve for the probit regression results from the second column of Table 11.3. The sample consists of 53,994 firm-year observations. The ROC curve shows the relation between the proportion of fraud detected and the proportion of false positives for all possible classification cut-points. The ROC curve is generated by taking each observation's estimated fraud probability, computing the sensitivity and false positives using that point as a cut-point, and then plotting the results.

both the percentage of fraud firm-years predicted and the percentage of the total dollar losses to fraud avoided are substantially lower. By comparison, the results for the subsample that excludes firms with prior publicly revealed fraud, shown in the fourth column, are very similar to the full sample. Thus, the difference in the last column is not due to some firms committing fraud numerous times. Rather, fraud is relatively easy to predict among firms with past regulatory and legal violations.

Out-of-Sample Prediction of Fraud

A key concern for any prediction model is out-of-sample validity. In this subsection, we test whether the within-sample predictions, reported in Panel B of Table 11.3, are robust out-of-sample. We do this in two ways: Panel C summarizes the out-of-sample predictive performance of each model in the post-2007 period. Panel D shows the results from K-fold cross-validation tests, which are explained in the next subsection.

The observations used in the prediction models reported in Panel A of Table 11.3 include only firm-years prior to August 1, 2007. To conduct an out-of-sample test, we search the SEC administrative proceedings and litigation releases and identify fraud cases that occurred between August 1, 2007 and July 31, 2010. Using Form ADV filings as of August 1, 2006, we assign each firm a predicted value based on the coefficients estimated within-sample. We then test whether these predicted values can accurately classify the out-of-sample fraud risk of the firms.

Panel C of Table 11.3 shows the proportion of fraud cases correctly predicted at a false positive rate of 5%. The proportion of fraud cases predicted out-of-sample is usually higher than within-sample, although given the small number of observations, this difference is not statistically significant. Also, although we use the within-sample cutoff values to classify the firms out-of-sample, the false positive rate does not increase.

K-Fold Cross-Validation Tests

As a further robustness test of the predictive models in Panel A of Table 11.3, we perform K-fold cross-validation tests over the period August 2001 through August 2007. The idea behind these tests is simple. Each model is estimated on a randomly selected subsample of firms, and the coefficient estimates from this subsample are used to classify the firms in the hold-out sample. Specifically, each firm in the sample is randomly assigned to one of ten groups (note that we randomly assign firms, and not firm-years, to avoid overstating the results due to non-independence). We then estimate the prediction model ten times, excluding each randomly formed group once. Each observation in the excluded group is assigned a predicted value, using the coefficients estimated from the observations in the other nine groups. The cutoff scores for fraud prediction are calculated within-sample and used to classify the observations in the hold-out sample. We repeat this process 20 times, for a total of 200 hold-out samples.

The results, shown in Panel D of Table 11.3, indicate that the predictive power of the models is only slightly lower in the hold-out samples. For example, the specification in the second column correctly predicts 150 fraud firm-years within-sample, compared to an average of 143.3 fraud firm-years in the hold-out samples. The K-fold test predicts a minimum of 135 and a maximum of 149 fraud firm-years across 20 repetitions, which suggests the model is quite stable.

The results of the out-of-sample and K-fold cross-validation tests support the robustness of the fraud predictions in Panel B. Note that these are robustness tests of the models' overall predictions and do not provide evidence as to the robustness of the individual coefficients. Overall, the results from the four panels of Table 11.3 show that the information investment advisers are required to disclose is relevant and useful for predicting fraud.

Annual Cross-Sectional Regressions

Although the models presented in Table 11.3 use observations from the entire sample period, which allows for relatively powerful tests, this may obscure time effects, which could arise in several ways. First, the actual rate of fraud could change over time due to changes in the legal or operating environment (e.g., poor performance could decrease the benefits of a reputation for honesty, thus increasing the incentive to commit fraud). Second, the detection rate could change over time. The median fraud persists for five and one-half years until detection. This suggests that the dependent variable for the 2006 cross-section likely includes fewer than half of the fraud cases that actually occurred in that year. By contrast, the 2001 cross-section likely includes a much higher proportion of the fraud cases that actually occurred in that year.

To examine whether there are time effects in the prediction of fraud, Panel A of Table 11.4 shows annual, cross-sectional probit regressions that predict investment fraud that occurs during the subsequent 12 months. For example, the model in column 1 uses Form ADV data available on August 1, 2001 to predict fraud that occurs from August 2001 through July 2002. Because fraud cases can persist for multiple years, these annual regressions are not independent, and aggregating coefficients across years could lead to faulty conclusions.

We test whether the coefficient estimates are significantly different across years with Wald tests. Because the same firm can appear in multiple years, we adjust the Wald tests for non-independence. The coefficients for Dedicated CCO and for Investment Company Act are significantly different across years at the 1% and 10% levels, respectively. Both variables are significant only in the early years of the sample. This change is partially due to the mutual fund late trading scandal that occurred in the early years of the sample. The firms involved in these cases managed funds that were registered under the Investment Company Act, and were mostly large financial conglomerates, which are more likely to have a dedicated CCO. Custody and Majority employee owned are also significantly different across years.

Note that because there are fewer observations in these annual cross-sectional regressions, the Wald tests have low power to reject the hypothesis that the coefficients are equal across years. For example, Referral fees is significant in Table 11.3, but in Table 11.4 Referral fees is significant in only two years. We cannot reject that the coefficients are jointly equal to zero, nor that they are jointly equal to the full-sample coefficient.

Panel B of Table 11.4 shows the ability of the cross-sectional regressions to predict fraud within-sample at a 5% false positive rate for each year. Chi-square tests show that the prediction rate of the annual cross-sectional regressions is significant in each year, which suggests the results in Table 11.3 are not driven by a single period. The proportion of fraud cases predicted, however, is lower in the last three years of the sample. This is partially due to the mutual fund late trading cases that occurred in the earlier years of the sample. Even after removing these cases, however, predictive accuracy appears to decline. This decline could indicate that the actual predictability of fraud has declined over time; although given the relatively low number of fraud cases in some years and the power of the annual tests, we cannot draw strong conclusions.

Initiation versus Continuance of Fraud

In the previous tests, the dependent variable does not distinguish between the initiation of a new fraud and the continuance of a preexistent fraud. In this section, by contrast, we test whether Form ADV data can be used to predict the initiation of a new fraud. This is important for three reasons. First, predicting fraud prior to initiation likely minimizes the harm. Second, ongoing fraud could affect the predictive variables, and thus, the previous tests blur the distinction between predicting future acts of fraud versus detecting ongoing fraud. Third, initiating a new fraud and continuing a preexistent fraud are economically different decisions. Lott (1996) shows that firms may initiate fraud in response to changes in their cost structure. Dechow, Sloan, and Sweeney (1996) and Dechow, Ge, Larson, and Sloan (2011) find evidence that companies initiate accounting fraud in response to corporate performance. Thus, certain predictive variables might measure a time-varying factor that triggers the initiation of fraud, whereas other variables might measure a time-invariant propensity toward fraud.

To address these issues, Panel A of Table 11.5 shows the results of a sequential logit regression with standard errors clustered by firm. The equation displayed in column 1 predicts the initiation of fraud in the subsequent year. The results in this column are qualitatively similar to the results in Table 11.3, and the chi-square test shows that the overall equation is statistically significant. The equation displayed in the second column predicts whether firms that have previously initiated a fraud will continue the fraud in the subsequent year. The insignificant chi-square test shows that the Form ADV variables have no incremental ability to distinguish which initiated fraud cases are continued into subsequent years. The insignificance of the second equation likely reflects that the Form ADV variables are quite stable over time and suggests

Table 11.5 **Initiation versus Continuance of Fraud**

The sample consists of 53,994 firm-year observations. The independent variables are taken from each firm's Form ADV filings as of August 1 each year from 2001 through 2006. Panel A shows the results of a sequential logit regression predicting fraud. The first column shows estimates of the probability that a firm initiates a fraud in the subsequent year. The second column shows estimates of the probability that a firm with a preexisting fraud continues that fraud into the subsequent year. Refer to Table 11.2 for variable definitions. In the interest of brevity, the constants are not reported. All significance tests are based on standard errors clustered by firm. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Panel B shows the proportion of initiated fraud cases that could be predicted within-sample.

Panel A: Predicting Initiation versus Continuance of Fraud

	Initiate	Continue
Past fraud	0.537 [1.31]	0.200 [0.33]
Past affiliated fraud	−0.417 [−1.33]	0.350 [0.65]
Past regulatory	0.707*** [3.87]	0.010 [0.03]
Past civil or criminal	0.447** [2.14]	−0.177 [−0.53]
Referral fees	0.240 [1.50]	−0.839*** [−2.78]
Interest in transaction	0.509*** [2.59]	0.153 [0.53]
Soft dollars	−0.124 [−0.70]	0.389 [1.62]
Broker in firm	0.349** [2.05]	−0.562** [−2.10]
Investment Company Act	0.612*** [3.04]	−0.023 [−0.07]
Custody	0.246 [1.33]	0.317 [1.13]
Dedicated CCO	−0.199 [−1.14]	0.305 [1.00]
Majority employee owned	−0.022 [−0.13]	−0.141 [−0.43]
Log (avg. acct. size)	−0.181*** [−4.63]	0.083 [1.10]
Percent client agents	0.009*** [3.56]	−0.003 [−0.51]
Hedge fund clients	0.034 [0.09]	−0.335 [−0.53]
Log (AUM)	0.155*** [4.78]	−0.068 [−1.15]
Log (firm age)	0.014 [0.34]	0.092 [1.50]
Model chi-square	199.7***	18.3

Panel B: Within-Sample Predictions

# Fraud Fraud predicted	87 33 37.9%	
# Clean firms Clean firm false positives	53,907 2,673 5.0%	

that these variables primarily measure a time-invariant component of the propensity for fraud.

Panel B of Table 11.5 shows the proportion of initiated fraud cases that could be predicted within-sample.[12] At a false positive rate of 5%, the model predicts 37.9% of initiated fraud cases, which suggests that the predictive ability of the model is not entirely due to continued fraud cases.

[12] We do not show prediction results for the proportion of preexistent fraud cases that are continued into the subsequent year. These predictions could be made only if preexistent fraud cases were readily observable, which is not the case.

Firm-Wide Fraud versus Fraud by Rogue Employees

In some cases, the executives of the firm commit or are aware of the fraud. In other cases, rogue employees evade their firms' internal control systems. A potential concern with the prior tests is that rogue employee fraud is likely more frequent at firms with many employees. If the predictive variables are correlated with the number of employees, this could lead to spurious correlations with rogue employee fraud. To ensure that this problem does not drive the results, we compare the predictability of firm-wide and rogue employee fraud.

Panel A of Table 11.6 shows the results of a multinomial probit regression. In the first column, the dependent variable equals one if the firm commits a firm-wide fraud in the subsequent year. In the second column, the dependent variable equals one if a rogue employee commits fraud in the subsequent year. The results for firm-wide fraud are very similar to the full-sample results, reported in column 3 of Table 11.3. At a false positive rate of 5%, the model can predict 24.2% of all firm-wide fraud cases, compared to 29.4% of all fraud cases. Panel B shows that, at a 5% false positive rate, the model can predict 74.6% of rogue employee fraud cases. Thus, although firm-wide fraud is more difficult to predict, the results in Table 11.3 are not entirely due to rogue employee fraud.

11.5 ARE INVESTORS COMPENSATED FOR FRAUD RISK?

Given that fraud is predictable, why would investors allocate money to firms with high fraud risk? One possibility is that the characteristics associated with high fraud risk may provide offsetting benefits that improve investment performance. Some investors could voluntarily accept high fraud risk in return for higher expected returns. Another possibility, which follows from the theoretical model of Lott (1996), is that investors pay a premium for advisers with low fraud risk, and pay lower fees to advisers with high fraud risk. In this section, we test whether investors are compensated for bearing fraud risk through higher returns or lower fees.

Until this point, the unit of observations has been firm-years. To test whether investors are compensated for fraud risk, however, requires fund data, which are only available for a subset of investment advisers. Hedge fund and institutional fund advisers are not required to report return data, which could create a selection bias if fraudulent advisers choose not to report returns. Thus, the results in this section are less general and may not apply to the full sample. Data are available for all mutual funds and mutual fund advisers, however, and so there is no selection bias for this category of funds.

We measure each firm-year observation's fraud risk as the predicted value from the probit regression in the third column of Table 11.3. In August of each year, we assign each fund its adviser's predicted fraud risk. We classify funds as high fraud risk if they are advised by a firm whose predicted fraud risk is greater than the 95th percentile of clean firms. This classification corresponds to the prediction cutoff used in Panel B of Table 11.3.

For each of the three return database samples, we form two equally weighted portfolios. In August of each year, we assign all funds classified as high fraud risk to one portfolio. The remaining funds are assigned to the low fraud risk portfolio. We then find the equally weighted returns of the portfolios for the subsequent 12 months.

See the Web Appendix for additional details and for results using value-weighted portfolios.

In Panel A of Table 11.7, we estimate alphas for the TASS hedge fund sample using the Fung and Hsieh (2001) eight-factor model.[13] For the CRSP mutual funds and PSN institutional funds, we estimate alphas using the Carhart (1997) model. For all three of the return database samples, we estimate alphas over the 72-month period from August 2001 to July 2007. Within all three return database samples, we do not find evidence that fraud risk is associated with higher risk-adjusted returns.

The regressions in Panel B of Table 11.7 test whether investors receive compensation for predicted fraud risk through lower fees. The control variables follow Cassar and Gerakos (2010) for hedge funds and Khorana, Servaes, and Tufano (2009) for mutual funds and institutional funds. All specifications control for fund style,[14] fund size, fund management company size, and year fixed effects. Within each of the three return database samples, we estimate pooled regressions with standard errors clustered by firm. We do not find evidence that high fraud risk is associated with lower fees in any of the samples.

The results in Panels A and B of Table 11.7 do not provide evidence that investors receive compensation for fraud risk. This finding has important implications for interpreting the prediction results. In Section 4.2, we noted that using a fraud prediction model requires an investor to avoid a considerable number of

[13] We are grateful to David Hsieh for providing the factors used for these regressions on his Web site: http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-FAC.xls

[14] We measure fund style using the Primary Category variable for the TASS hedge funds, Lipper categories for the CRSP mutual funds, and reported market capitalization and value/growth categories for the PSN funds.

Table 11.6 Firm-Wide versus Rogue Employee Fraud

The sample consists of 53,994 firm-year observations. The independent variables are taken from each firm's Form ADV filings as of August 1 each year from 2001 to 2006. Panel A shows the results of a multinomial probit regression predicting fraud. In the first column, the dependent variable equals one for firms that experience a firm-wide fraud in the subsequent year. In the second column, the dependent variable equals one for firms that experience a rogue employee fraud in the subsequent year. The excluded category is clean firms. Refer to Table 11.2 for variable definitions. In the interest of brevity, the constants are not reported. All significance tests are based on standard errors clustered by firm. *Z*-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The columns in Panel B correspond to the columns in Panel A. Panel B shows the proportion of fraud that could be predicted within-sample.

Panel A: Predicting Firm-Wide versus Rogue Employee Fraud

	Firm	Rogue
Past fraud	−0.698	1.107**
	[1.00]	[2.03]
Past affiliated fraud	−0.700*	0.105
	[1.72]	[0.18]
Past regulatory	0.697***	0.888**
	[3.45]	[2.09]
Past civil or criminal	0.066	1.600***
	[0.25]	[3.21]
Referral fees	0.194	0.926
	[1.15]	[1.59]
Interest in transaction	0.435**	1.185*
	[2.10]	[1.69]
Soft dollars	−0.03	−0.925*
	[0.16]	[1.88]
Broker in firm	0.327*	0.88
	[1.85]	[0.94]
Investment Company Act	0.656***	0.351
	[3.03]	[0.74]

Table 11.6 Continued

Panel A: Predicting Firm-Wide versus Rogue Employee Fraud

	Firm	Rogue
Custody	0.340*	−0.729
	[1.73]	[1.49]
Dedicated CCO	−0.312	0.301
	[1.59]	[0.67]
Majority employee owned	−0.013	0.026
	[0.07]	[0.05]
Log (avg. acct. size)	0.144***	−0.294***
	[3.25]	[3.27]
Percent client agents	0.009***	0.003
	[3.45]	[0.41]
Hedge fund clients	0.001	−0.441
	[0.00]	[0.40]
Log (AUM)	0.121***	0.372**
	[3.27]	[2.45]
Log (firm age)	0.005	0.176
	[0.12]	[0.93]
Model chi-square	144.3***	193.8***

Panel B: Within-Sample Predictions

	Firm	Rogue
# Fraud	450	67
Fraud predicted	109	50
	24.2%	74.6%
# Clean firms		53,477
Clean firm false positives		2,673
		5.0%

honest advisers (false positives). The results in this section suggest that there is little cost to avoiding these false positives.

The lack of compensation for fraud risk does not necessarily imply that investors are irrational. First, investors could be compensated in some other, unmeasured fashion. Second, our measure of predicted fraud risk uses Form ADV disclosures. Investors may predict fraud risk based on reputations, personal contacts, or other information; investors could be compensated

Table 11.7 Fraud Risk, Alphas, and Fees

In this table we test the relation of fraud risk with alphas and fees. We merge the Form ADV sample with the TASS hedge fund (TASS), CRSP mutual fund (CRSP), and PSN institutional fund (PSN) databases. Each investment adviser's fraud risk is defined as the predicted value from the regression reported in column 3 of Table 11.3. We assign this measure of fraud risk to each fund managed by that investment adviser. For each of the return databases, we create two equally weighted portfolios based on the funds' predicted fraud risk. The high fraud risk portfolio includes all funds advised by firms whose predicted fraud risk is greater than the 95th percentile of clean firms. The low fraud risk portfolio includes all other funds. We estimate alphas using monthly returns for each portfolio. We use the Fung and Hsieh (2001) model for the TASS sample and the Carhart (1997) model for the CRSP and PSN samples. For the CRSP mutual fund and PSN Informa databases, we include only equity funds. High-Low is the alpha of a portfolio long high fraud risk funds and short low fraud risk funds. The *t*-statistics, reported in square brackets, are adjusted using the method of Newey and West (1987) with three lags. Panel B reports the relation between fraud risk and fees. The dependent variables for the TASS sample are the management and incentive fees. The dependent variable for the CRSP sample is expense ratios. The dependent variable for the PSN sample is the reported fee percentage charged on a $50 million account. High risk equals one if the fund is advised by a firm whose predicted value from column 3 of Table 11.3 is greater than the 95th percentile of clean firms. Log(fund AUM) is the logarithm of assets under management for the fund. Log(fund age) is the logarithm of fund age in years. Fund offshore equals one if the fund is registered offshore (i.e., non-US). Leverage equals one if the fund uses leverage. Log(firm AUM) is the logarithm of assets under management for the firm to which the fund belongs. Turnover % is the annual percentage turnover of the fund's portfolio. Index fund equals one if the fund is an index fund. Style fixed effects are created using Primary Category from TASS, Lipper Objective in CRSP, and Investment Style and Market Cap from PSN. Year dummies and constants are included but not reported. The variables are sampled annually as of August 1 of each year from 2001 to 2006. The standard errors are clustered by firm. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Portfolio Alphas			
	Low Fraud Risk	**High Fraud Risk**	**High-Low**
TASS	0.0046***	0.0034***	−0.0012**
	[5.96]	[4.24]	[2.11]
CRSP	−0.0021***	−0.0021***	−0.00002
	[5.52]	[4.43]	[0.13]
PSN	0.0001	0.0001	0.00005
	[0.14]	[0.20]	[0.19]

Panel B: Fees				
	TASS		**CRSP**	**PSN**
	Management	**Incentive**	**Expense Ratio**	**Management**
High risk	−0.089	0.022	0.047	−0.004
	[1.32]	[0.04]	[1.40]	[0.25]
Log (fund AUM)	0.008	−0.017	−0.077***	−0.0004
	[0.88]	[0.22]	[10.21]	[0.14]
Log (fund age)	−0.055***	−0.221	0.045***	
	[3.72]	[1.56]	[2.76]	
Fund offshore	0.115***	0.042		
	[3.54]	[0.15]		

Table 11.7 Continued

Panel B: Fees				
	TASS		CRSP	PSN
	Management	Incentive	Expense Ratio	Management
Leverage	0.050* [1.76]	1.818*** [5.04]		
Log (firm AUM)			−0.024*** [3.07]	−0.004 [1.07]
Turnover %			0.0005*** [4.02]	0.0002** [2.31]
Index fund			−0.740*** [11.80]	
Style fixed effects	Yes		Yes	Yes
Year fixed effects	Yes		Yes	Yes
Observations	5,998		20,767	8,682
Adjusted R^2	0.141		0.233	0.227

for some alternative measure of fraud risk. Finally, because of barriers to accessing and using the Form ADV disclosures, investors might perceive the costs of estimating fraud risk to outweigh the benefits.

11.6 DATA ACCESS AND IMPLEMENTATION

In this section, for each year in the sample, we test how well investors could have implemented predictive models using only Form ADV data that had previously been publicly accessible. These tests differ from the predictive regressions in Table 11.3, which use information from the full-sample period and thus, do not directly address how well an investor could have predicted fraud during the sample (e.g., the fraud predictions in 2003 are based on coefficients estimated using data from 2001 to 2006).

During the sample period, the SEC did not provide public access to historical Form ADV filings; investors could access only a contemporaneous cross-section. For this reason, we compare two types of predictive models. In the first, we estimate predictive models that use only the contemporaneous cross-section of Form ADV filings. These tests mimic the predictions an investor could have made during the sample period, given the actual data access policies in place. In the second, we estimate predictive models that use data from an annual panel of historical Form ADV filings. These tests mimic the predictions an investor could have made if historical filings had been publicly accessible.

Table 11.8 shows the results of fraud prediction models that use only the contemporaneously accessible cross-section of Form ADV filings as of August 1 of each year. To illustrate, in the column labeled Aug '05, the independent variables are taken from each firm's most recent Form ADV filing as of August 1, 2005. The dependent variable equals one for all firms with an observable prior fraud case (i.e., a fraud case that occurred between September 28, 1995 and July 31, 2005, and which was identified in an SEC administrative proceeding or litigation release filed before July 31, 2005). We use the coefficient estimates from this regression to make out-of-sample predictions of the fraud cases that occur between August 1, 2005 and July 31, 2006.

For comparison purposes, Table 11.9 shows the results of fraud prediction models that use an annual panel of historical Form ADV filings. Like Table 11.8, these regressions use only information that existed at the time of the prediction. But unlike Table 11.8, they use information that was not contemporaneously accessible by the public. To illustrate, in the column labeled August 2005, the

independent variables are taken from each firm's Form ADV filings as of August 1 in 2001–2004. For each August 1 firm-year observation, the dependent variable equals one if the firm commits fraud during the subsequent 12 months, and the fraud is publicly revealed before August 1, 2005. We combine the coefficient estimates from this model with each firm's Form ADV data as of August 1, 2005 to make out-of-sample predictions of the fraud cases that occur between August 1, 2005 and August 1, 2006.

The models presented in Tables 11.8 and 11.9 differ in several ways. Most obviously, the panel models in Table 11.9 use more data to estimate fraud risk than do the cross-sectional models in Table 11.8. More important, the models in Table 11.8 are backward-looking: they show the relation between contemporaneous variables and past fraud. If the contemporaneous Form ADV filings are all that is publicly accessible, then investors can only estimate fraud risk from backward-looking regressions (e.g., Brown, Goetzmann, Liang, and Schwarz, 2008). If historical filings are accessible, investors can estimate forward-looking prediction models and then estimate fraud risk by combining the estimated coefficients with the contemporaneous disclosures of the firms. This is a conceptually important distinction. The backward-looking regressions only include the subsample of firms that survived the legal consequences

Table 11.8 Point-In–Time Tests Using Publicly Accessible Data

Each column uses only the contemporaneously accessible cross-section of Form ADV filings as of August 1 of that year. Panel A shows the estimates from cross-sectional probit regressions. The dependent variable equals one for firms which have a publicly observed prior history of fraud (fraud occurred and was detected between January 1, 1996 and August 1 of the year in which the independent variables are observed). The independent variables reflect the publicly accessible data as of August 1 of the year in the column. Refer to Table 11.2 for variable definitions. In the interest of brevity, we do not report coefficients for the constants. Standard errors are robust. *Z*-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Panel B shows the proportion of fraud that could be predicted in year $T + 1$, using Form ADV data at time T as inputs to the prediction model in the aligned column of Panel A.

Panel A: Point-in-Time Cross-Sections						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
Past affiliated fraud	0.757***	0.774***	0.111	−0.006	−0.067	−0.188
	[2.70]	[3.14]	[0.39]	[0.03]	[0.30]	[0.87]
Past regulatory	0.705***	0.976***	0.984***	0.820***	0.953***	1.166***
	[3.99]	[4.66]	[6.56]	[5.24]	[7.40]	[8.69]
Past civil or criminal	0.642***	0.445**	0.594***	0.620***	0.409**	0.464***
	[2.85]	[2.28]	[3.44]	[3.59]	[2.53]	[3.50]
Referral fees	0.437**	0.347*	0.333**	0.234	0.137	0.127
	[2.17]	[1.91]	[2.14]	[1.64]	[1.15]	[1.07]
Interest in transaction	0.366	0.447**	0.019	−0.251	0.063	0.081
	[1.56]	[2.10]	[0.11]	[1.39]	[0.43]	[0.65]
Soft dollars	−0.401**	−0.311*	−0.210	−0.342**	−0.102	0.017
	[2.14]	[1.91]	[1.32]	[2.26]	[0.85]	[0.14]
Broker in firm	0.093	0.045	−0.090	−0.209*	−0.157	−0.133
	[0.54]	[0.25]	[0.69]	[1.71]	[1.26]	[1.05]
Investment Co. Act	0.206	0.155	−0.088	0.284*	0.118	0.143
	[0.88]	[0.71]	[0.43]	[1.65]	[0.66]	[0.87]

Table 11.8 Continued

Panel A: Point-in-Time Cross-Sections						
	Aug '01	**Aug '02**	**Aug '03**	**Aug '04**	**Aug '05**	**Aug '06**
Custody	−0.092	−0.019	−0.039	0.165	−0.067	−0.007
	[0.48]	[0.11]	[0.23]	[1.11]	[0.55]	[0.06]
Dedicated CCO	−0.321	−0.178	−0.035	0.174	0.089	0.098
	[0.86]	[0.78]	[0.15]	[1.05]	[0.78]	[0.92]
Majority emp. owned	−0.114	0.102	0.011	−0.056	0.083	0.148
	[0.51]	[0.55]	[0.06]	[0.38]	[0.67]	[1.16]
Log (avg. acct. size)	−0.161***	0.114***	−0.044	−0.062*	−0.069**	−0.067**
	[3.52]	[3.09]	[1.09]	[1.78]	[2.10]	[2.13]
Percent client agents	0.000	−0.001	−0.001	0.002	0.001	0.000
	[0.12]	[0.22]	[0.26]	[0.86]	[0.28]	[0.06]
Hedge fund clients	0.384	0.074	−0.079	−0.122		−0.421
	[0.93]	[0.20]	[0.20]	[0.32]		[1.08]
Log (AUM)	0.121***	0.095***	0.037	0.057*	0.057**	0.048*
	[3.39]	[2.86]	[1.20]	[1.92]	[2.07]	[1.85]
Log (firm age)	0.056*	0.064	0.183***	0.160***	0.124**	0.197***
	[1.70]	[1.30]	[3.29]	[2.79]	[2.20]	[3.99]
Model chi-square	224.8***	116.2***	115.1***	150.1***	139.4***	155.7***
Observations	7,352	7,747	8,562	9,088	9,110	10,383
Panel B: Out-of-Sample Predictions						
# Fraud	104	116	115	83	59	40
Fraud predicted	28	33	27	20	12	10
	26.9%	28.4	23.5	24.1	20.3	25.0
# Clean firms	7,248	7,631	8,447	9,005	10,803	10,343
Clean firm false positives	360	374	419	456	470	552
	5.0%	4.9	5.0	5.1	4.4	5.3

of committing fraud, which could bias the coefficients.[15] The forward-looking models in Table 11.9 have one disadvantage, however: These models require at least two years of data to estimate, and so it is not possible to estimate this model for the year beginning August 1, 2001.

For the regressions reported in Tables 11.8 and 11.9, our main interest is not the coefficient estimates, but rather the models' ability to predict fraud. Specifically, we are interested in two issues. First, could the prediction results in Table 11.3 have been

[15] Nearly one-third of the firms at which fraud is detected cease operations within one year. Duration models, presented in the Web Appendix, show a strong relation between the detection of fraud and firm closure.

Table 11.9 **Predictions Using a Panel of All Prior Years**

Each column represents the predictions an investor could have made at a specified point-in-time had historical Form ADV data been publicly available. For each column, the sample consists of a panel of all previously available annual Form ADV filings as of August 1 of each year. (E.g., in August 2002, the independent variables are taken from the August 2001 cross-section of Form ADV filings. In August 2003, the independent variables are taken from the August 2002 and August 2001 samples of Form ADV filings.) Panel A shows the results of fraud prediction models that use all prior Form ADV filings to predict fraud. For each firm-year observation, the dependent variable equals one if the firm commits a fraud during the subsequent 12 months. Refer to Table 11.2 for variable definitions. In the interest of brevity, we do not report coefficients for the constants. Standard errors are clustered by firm and year. *Z*-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Panel B shows the proportion of fraud that could be predicted in year $T + 1$, using Form ADV data at time T as inputs to the prediction model in the aligned column of Panel A.

Panel A: Panel of All Prior Years					
	Aug '02	**Aug '03**	**Aug '04**	**Aug '05**	**Aug '06**
Past fraud	0.280	0.226	0.102	0.178	0.254
	[0.82]	[0.94]	[0.37]	[0.79]	[1.21]
Past affiliated fraud	−0.321	−0.271*	−0.237*	−0.219*	−0.187
	[1.23]	[1.83]	[1.91]	[1.89]	[1.52]
Past regulatory	0.187*	0.213***	0.261***	0.288***	0.293***
	[1.89]	[3.53]	[3.42]	[3.82]	[4.15]
Past civil or criminal	0.239*	0.220***	0.176*	0.211**	0.181*
	[1.73]	[2.61]	[1.80]	[2.25]	[1.96]
Referral fees	0.041	0.031	0.045	0.061	0.084
	[0.45]	[0.67]	[0.83]	[1.15]	[1.53]
Interest in transaction	0.265***	0.257***	0.250***	0.227***	0.216***
	[2.78]	[4.92]	[4.27]	[3.42]	[3.29]
Soft dollars	−0.037	−0.053	−0.041	−0.034	−0.033
	[0.40]	[1.04]	[0.74]	[0.61]	[0.61]
Broker in firm	0.202**	0.162***	0.127**	0.117*	0.111*
	[2.33]	[2.86]	[2.04]	[1.93]	[1.89]
Investment Co. Act	0.245**	0.289***	0.301***	0.304***	0.276***
	[2.43]	[4.17]	[4.60]	[4.68]	[3.59]
Custody	0.006	0.048	0.094	0.075	0.088
	[0.06]	[0.72]	[1.20]	[1.07]	[1.37]
Dedicated CCO	0.247	0.305***	0.346***	0.199	0.027
	[1.53]	[3.10]	[3.46]	[1.29]	[0.22]
Majority emp. owned	−0.089	−0.099*	−0.053	−0.005	0.016
	[0.88]	[1.90]	[0.71]	[0.06]	[0.19]

Table 11.9 Continued

Panel A: Panel of All Prior Years					
	Aug '02	**Aug '03**	**Aug '04**	**Aug '05**	**Aug '06**
Log (avg. acct. size)	−0.100***	−0.094***	−0.087***	−0.079***	−0.072***
	[3.90]	[7.75]	[5.85]	[4.68]	[4.01]
Percent client agents	0.004***	0.004***	0.004***	0.003***	0.003***
	[3.03]	[4.18]	[4.29]	[3.90]	[3.94]
Hedge fund clients	0.006	0.072	0.114	0.107	0.053
	[0.02]	[0.52]	[0.87]	[0.83]	[0.42]
Log (AUM)	0.091***	0.084***	0.076***	0.067***	0.061***
	[4.24]	[7.72]	[5.46]	[4.39]	[3.89]
Log (firm age)	0.019	0.015	0.008	0.008	0.007
	[1.00]	[1.44]	[0.68]	[0.71]	[0.69]
Model chi-square	138.0***	543.8***	304.8***	218.7***	205.5***
Observations	7,352	15,099	23,661	32,749	43,611
Panel B: Out-of-Sample Predictions					
# Fraud	116	115	83	59	40
Fraud predicted	48	38	21	12	10
	41.4%	33.0	25.3	20.3	25.0
# Clean firms	7,631	8,447	9,005	10,803	10,343
Clean firm false positives	386	402	496	603	517
	5.1%	4.8	5.5	5.6	5.0

achieved during the sample. Second, we are interested in comparing the predictive ability of the models in Tables 11. 8 and 11.9. Between August 1, 2002 and August 1, 2007, there were a total of 413 fraud firm-years. At a false positive rate of 5%, the cross-sectional regressions, shown in Table 11.8, predict 24.7% of the fraud cases (the fraud firm-years predicted between August 2002 and August 2007 sum to 102). The panel regressions using all prior years, shown in Table 11.9, predict 31.2% of the fraud firm-years (a sum of 129 fraud firm-years). A chi-square test of classification accuracy shows that the panel regressions in Table 11.9 predict a significantly larger number of fraud cases (p-value < 0:01). Although the absolute difference in predictive accuracy is only 6.5% points (an improvement of 26.3% relative to the model in Table 11.8), these tests provide evidence that public access to historical Form ADV filings could benefit investors. Moreover, the marginal cost to the SEC of allowing public access would be quite low.

To implement a fraud prediction model, such as those tested in this paper, an investor would have had to collect manually a large number of Form ADV filings, convert the filings into a database, and estimate a prediction model. For most investors, the cost of individually downloading thousands of Form ADV filings may well have exceeded the perceived benefits. This problem is exacerbated by the fact that investors are atomistic: Even if the aggregate benefit of processing the disclosed information outweighs the cost to a single investor, the benefit to any single investor may be insufficient. As shown by Becker (1968), the socially optimal level of a crime occurs when the marginal benefit from a further reduction in the

crime is equal to the marginal cost of enforcement. Allowing public access to historical Form ADV filings would reduce the marginal cost of increased enforcement by facilitating investors' use of these data. This, in turn, should reduce the marginal benefit to an investment adviser of committing fraud due to an increase in the probability of detection. Thus, improved public access to these disclosure data could reduce the occurrence of fraud.

11.7 CONCLUSION

This paper finds that required disclosures related to past regulatory and legal violations, conflicts of interest, and monitoring are significant predictors of investment fraud. We stress, however, that prediction does not imply a causal relation between the disclosed information and fraud. Although fraud is predictable, we do not find evidence that investors are compensated for fraud risk. To explain this puzzling fact, we examine the barriers to implementing fraud prediction models.

If, during the period August 2001–2007, investors had avoided the 5% of firms with the highest ex ante predicted fraud risk, they could have avoided more than $4 billion in losses from fraud. Based on the SEC's estimate of 9.01 hours to fill out Form ADV and an assumed cost of $1,000 per hour, during this same period the direct costs of disclosure were at most $500 million. Thus, even ignoring the deterrent effect of disclosure, this simple, back-of-the-envelope calculation suggests that the benefits of Form ADV substantially outweigh the costs. During the sample period, the investing public's ability to develop and use predictive models based on Form ADV data was potentially limited because the SEC did not provide access to historical data. As a result, the realized benefits of disclosure during the period may have been lower. The results suggest that improving public access to comprehensive historical disclosures could increase the benefits these disclosures were meant to provide.

APPENDIX A

Variable Definitions

Variable	Definition	Data Source
Past fraud	The firm committed a previously detected fraud	SEC administrative proceeding or litigation release was filed for firm prior to August 1 of firm-year observation
Past affiliated fraud	An affiliate of the firm committed a previously detected fraud	SEC administrative proceeding or litigation release was filed for affiliated firm prior to August 1 of firm-year observation and Form ADV Schedule D Section 7.A reports fraud firm as affiliate
Past regulatory	Filed a regulatory disclosure reporting page (DRP)	One of more of: Items 11c1–3, 11d1–5, 11e–4
Past civil or criminal	Filed a criminal or civil DRP	One of more of: Items 11a1–2, 11b1–2, 11h1a, 11h1b, 11h1c, 11h2
Referral fees	Do you or any related person, directly or indirectly, compensate any person for client referrals?	Item 8f
Interest in transaction	Do you or any related person: buy (or sell) securities from advisory clients; recommend securities in which you have an ownership interest or serve as underwriter, general or managing partner or have any other sales interest?	One of more of: Items 8a1, 8a3, 8b2, 8b3
Soft dollars	Do you or any related person receive research or benefits other than execution from a broker-dealer or a third party in connection with client securities transactions?	Item 8e

Variable	Definition	Data Source
Broker in firm	Employs registered representatives of a broker-dealer	Item 5b2 > 0
Investment Company Act	Investment adviser (or sub-adviser) to an investment company registered under the Investment Company Act	Item 2a4
Custody	Do you or any related person have custody of any advisory clients' cash or securities?	One of more of: Items 9a1–2, 9b1–2
Dedicated CCO	CCO has no other stated role within firm	CCO on Schedule A has no other "Title or Status"
Majority employee owned	Over 50% aggregate employee ownership	Imputed using Dimmock, Gerken,& Marietta-Westberg (2011) method
Log (avg. acct. size)	Logarithm of assets under management per client	Log (Item 5f2c/(Item 5f2f + 1) + 1)
Percent client agents	Percent of banking, mutual, pension, charitable, corporate, and government clients	Sum of items: 5d3, 5d4, 5d5, 5d7, 5d8, 5d9 imputed using Dimmock, Gerken, & Marietta-Westberg (2011) method
Hedge fund clients	Primarily hedge fund clients	Item 5d6 ≥ 75%
Log (AUM)	Logarithm of assets under management	Log (Item 5f2 + 1)
Log (firm age)	Logarithm of firm age in years	Log (years since date registered with the SEC)

APPENDIX B

Supplementary Data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfineco.2012.01.002.

References

Becker, G., 1968. Crime and punishment: an economic approach. Journal of Political Economy 76, 169–217.

Bollen, N., Pool, V., 2008. Conditional return smoothing in the hedge fund industry. Journal of Financial and Quantitative Analysis 43, 267–298.

Bollen, N., Pool, V., 2009. Do hedge fund managers misreport returns? Evidence from the pooled distribution. Journal of Finance 64, 2257–2288.

Bollen, N., Pool, V., 2010. Predicting Hedge Fund Fraud with Performance Flags. Unpublished Working Paper. Vanderbilt University and Indiana University.

Brown, S., Goetzmann, W., Liang, B., Schwarz, C., 2008. Mandatory disclosure and operational risk: evidence from hedge fund registration. Journal of Finance 63, 2785–2815.

Brown, S., Goetzmann, W., Liang, B., Schwarz, C., 2009. Estimating operational risk for hedge funds: the w-score. Financial Analysts Journal 65, 43–53.

Busse, J., Goyal, A., Wahal, S., 2010. Performance persistence in institutional investment management. Journal of Finance 65, 765–790.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Cassar, G., Gerakos, J., 2010. Determinants of hedge fund internal controls and fees. Accounting Review 85, 1887–1919.

Darby, M., Karni, E., 1973. Free competition and the optimal amount of fraud. Journal of Law and Economics 16, 67–88.

Dechow, P., Ge, W., Larson, C., Sloan, R., 2011. Predicting material accounting misstatements. Contemporary Accounting Research 28, 17–82.

Dechow, P., Sloan, R., Sweeney, A., 1996. Causes and consequences of earnings manipulation: an analysis of firms subject to enforcement actions by the SEC. Contemporary Accounting Research 13, 1–36.

Dimmock, S., Gerken, W., Marietta-Westberg, J., 2011. What Determines the Allocation of Managerial Ownership Within Firms? Unpublished working paper. Nanyang Technological

University, Auburn University, and U.S. Securities and Exchange Commission.

Fung, W., Hsieh, D., 2001. The risk in hedge fund strategies: theory and evidence from trend followers. Review of Financial Studies 14, 313–341.

Karpoff, J., Lott, J., 1993. The reputational penalty firms bear from committing criminal fraud. Journal of Law and Economics 36, 757–802.

Karpoff, J., Lott, J., Wehrly, E., 2005. The reputational penalties for environmental violations: empirical evidence. Journal of Law and Economics 68, 653–675.

Khorana, A., Servaes, H., Tufano, P., 2009. Mutual fund fees around the world. Review of Financial Studies 22, 1279–1310.

Klein, B., Leffler, K., 1981. The role of market forces in assuring contractual performance. Journal of Political Economy 89, 615–641.

Lott, J., 1996. The level of optimal fines to prevent fraud when reputations exist and penalty clauses are unenforceable. Managerial and Decision Economics 17, 363–380.

Murphy, D., Shrieves, R., Tibbs, S., 2009. Understanding the penalties associated with corporate misconduct: an empirical examination of earnings and risk. Journal of Financial and Quantitative Analysis 44, 55–83.

Newey, W., West, K., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Straumann, D., 2009. Measuring the quality of hedge fund data. Journal of Alternative Investments 12, 26–40.

Zitzewitz, E., 2006. How widespread was late trading in mutual funds? American Economic Review 96, 284–289

INDEX

C

D

E

F

G

H

I

J

K

L

M

Q

R

S

T

U

V

W

X

Y

Z